Development of Stage – Distance – Discharge Relationship and Rating Curve using Least Square Method

Nariman Yahya Othman a*, Zahra Abd Saleh a, Zainab Ali Omran a

a University of Babylon, College of Engineering, Department of Civil Engineering, Babylon, Iraq.

Received 14 April 2019; Accepted 01 August 2019

Abstract

For any river, besides the importance of stage – discharge relationship (rating curve), a stage-distance-discharge relationship is of more significance. The accurate estimation of both relationships along a river reach is considered a key point for various applications of water resources engineering such as operation and management of water resources projects, designing of hydraulic structures, and sediment analysis. In this paper, both relationships were established for the Shatt Al – Hillah river reach by applying multiple linear regression and simple linear regression using least square method for determining regression equations. Twelve gauging stations including three primary and nine secondary stations were considered for this method. Moreover, for evaluating the performance of both regressions, statistical measures such as coefficient of determination, root mean square error, mean square error, and Thiel's factor were used. The study results generally indicate a superior performance of both modeling techniques. MLR model was able to predict and mimic the stage-distance-discharge relationship with a coefficient correlation of about 0.932, while SLR model was able to predict three rating curves for the three primary stations with coefficient correlation of about 0.960, 0.943, and 0.924 respectively.

Keywords: MLR; Multiple Linear Regression; Discharge; Water Elevation; Least Squares Method; SLR; Simple Linear Regression.

1. Introduction

A vital mater for variation of hydrologic requests like water resources and planning, hydraulic, hydrologic modeling, design of structures water conveyance and sediment analyses, etc., depends on the information about flow discharge in rivers. However, it is commonly known that the collection of direct measurements of flow discharge is costly and challenging. Therefore, it has been a corporate practice to exchange records of elevations which are less expensive and easier to measure into discharges by using a pre-established (elevation, discharge) correlation, though this can only be applied for a specific cross-sectional area of the river and not along it. Therefore, established water elevation-discharge-distance relationship is considered an important issue for dependable design, management, and planning of water resources projects. The development of such relationship involved two steps. In the first step the collection of the field data for stage (water elevation (E)), discharge (Q) at specified distance of the river (X) should be prepared. Then, in the second step a suitable method should be selected for establishing a mathematical model which associates the previous parameters together, i.e., (E, X and Q). In this study, multiple linear regression (MLR) for modeling the relationship between two or more variables was used as indicated by Freedman (2005) [1], Cook and Weisberg (1982) [2], and Rencher and Christensen (2012) [3]. The general form of the Equation of the MLR is:

\[ \hat{Y} = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \]  

(1)
The slopes of beta predicated factors, and ε is residuals between observed and predicted dependent variable.

In this work, Ŷ was the estimated water elevation (E), x1 was the distance (X) along the river, and x2 was the discharge (Q). Then field data of the water elevation, longitudinal distance, and discharge were used to obtain values of αo, β1, and β2 which predicts the equation of calculation of water elevation depending on distance and discharge. Multiple Linear Regression method is commonly used regression method in different engineering applications. Li and Wang (2019) considered the deformation of concrete dam under loads as a case study to make a comparison between Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models [4]. Li et al. (2017) proposed Multiple Linear Regression analysis for developing the Yellow river diversion and concluded the importance of the main factors effecting on the diversion of the river [5]. Patel et al. (2016) illustrated the rainfall – runoff relations using the Multiple Linear Regression as a based technique [6]. Seeboonruang (2017) established a (MLR) model to predict an equation representing salinity in shallow groundwater [7]. Ghimire and Reddy (2010) used different algorithms methods for developing a stage discharge rating curve in river and one of these methods was multiple linear regression (MLR) [8]. Knochenmus and Yobbi (2001) used regression techniques as multiple linear regression and least square regression to predict calculation to simulate the relation between groundwater levels and springs flows [9]. Finally, Al-Mukhtar and Al-Yaseen (2019) used multiple linear regression model as one of three different models to predict a water quality model for a Marsh in Iraq [10]. Measurement of water flow is a significant feature of hydrology associated project, which is as important as water quality monitoring, geomorphology, and flooding [11-13]. The rating curve is an essential technique, being very vital in discharge calculation [14, 15]. Also, for various hydrological applications such as irrigation operation, water resources management, and sediment controlling, hydrologic modeling available of accurate rating curve is very important [16]. Discharge, in addition of its dependence on stage, depends on cross-sectional area, bed slope of river, bed roughness, etc. the relations between discharge and elevation for any river is also very important; such relationships are often referred to as a rating curve and for the elevation – discharge relation (rating curve) for the three primary gage stations. If E represents Elevation for discharge Q, then the relationship between E and Q can possibly be approximated using the equation given by Herschy (1999) [17] and Kennedy (1984) [18]:

\[ Q = \alpha (E - E_o)^\eta \]  

Where \( \alpha \) and \( \eta \) are rating curve constants, and \( E_o \) is a constant representing the gauge reading which corresponds to zero discharge. The constant \( E_o \) is a hypothetical parameter which cannot be measured in the field therefore its value was taken from BWRD (2015) for each gaging station [19]. By using field data and applying least squares method, values of \( \alpha \) and \( \eta \) for the gaging stations can be determined. Many researchers have established stage – discharge relationship (rating curve) using different techniques. Muzzammil et al. (2018) used Excel solver technique for developing a rating curve [20]. Alfa et al. (2018) established a rating curve for Ofu River in Nigeria depending on linear regression analysis and by using analysis tool of Microsoft Excel 2007 [21]. While, Kavousizadeh (2019) introduced a determination of rating curve in compound open channels [22]. Verification for both developed equations [(stage (elevation)-discharge-distance) equation and (elevation-discharge) equation] has been measured by the determination of \( R^2 \) factor, Mean Square Error (MSE), Root Mean Square Error (RMSE), and Thiel's factor (U).

2. Describing Al-Hillah River

Al-Hillah River is the main river in Hillah City in Iraq (Figure 1), which provides all the demanded water for the city, irrigating about 67890 km² of agricultural lands. The river length from (Hindiya barrage upstream the river to Dora regulator downstream the river) is 51.100 km. The minimum and the maximum discharge at upstream of the river are [(50 and 230) m³/s] respectively. Whereas, the minimum and the maximum elevation at upstream of the river are [(29.7 and 31.30) (meter above sea level) m.a.s.l.] respectively. The river has ten intakes branches with their totals discharges of (42.585 m³/s) (BWRD, 2015) [19].
3. Methodology

3.1. Determining the Stage – Discharge – Distance Relationship by Multiple Linear Regression

As mentioned previously, the water elevation (E), discharge (Q), and distance (X) relation were obtained by using Multiple Linear Regression (MLR); where Equation 1 tends to be:

\[ \hat{E} = \alpha_0 + \beta_1 X + \beta_2 Q \]  

(3)

Equation 3 is known as sample regression equation [23]. For n observations and using Least Square Method (LSM), a set of normal equations would be predicted (Ezekiel and Fox (1959)) [24]. For two independent variables, as in the case study, the normal equations are:

\[ \sum_{i=1}^{n} X_i^2 \times \beta_1 + \sum_{i=1}^{n} (X_i \times Q) \times \beta_2 = \sum_{i=1}^{n} E_i \times X_i \]  

(4)

\[ \sum_{i=1}^{n} (X_i \times Q_i) \times \beta_1 + \sum_{i=1}^{n} Q_i^2 \times \beta_2 = \sum_{i=1}^{n} E_i \times Q_i \]  

(5)

\[ \alpha_0 = \hat{E} - \beta_1 \times \bar{X} - \beta_2 \times \bar{Q} \]  

(6)

Where: \( \alpha_0 \) is the constant of the linear regression equation and \( \beta_1 \) and \( \beta_2 \) are the beta predicated factors and they are calculated by solving Equations 4 to 6.

3.2 Determining the stage – discharge relation by Simple Linear Regression

The relation between (E and Q) was represented by Equation 2 given in section 1. By calculating \( \alpha \) and \( \eta \), the best fit curve for n observations of stage and discharge were obtained. By taking natural logarithms for Equation 2, the equation is:

\[ \ln Q = \ln (\alpha \ (E - E_0)^\eta) \]  

(7)

And as a linear form:

\[ Y = a \times x + b \]  

(8)

Where, \( Y = \ln Q \), \( a = \alpha \), \( x = \ln (E - E_0) \), and \( b = \ln \eta \), then as Least Squares method (LSM) for n observations:
After obtaining $\alpha$ and $\eta$ the sum of the residuals, $(\sum(Q_{obs} - Q_{pre})^2)$ was calculated whenever the residuals is minimum, which means the best model to represent rating curve was obtained for the desired gauging station [25, 26].

3.3. Determining the Goodness of Fit

The performance of all the equations predicted was evaluated using $R^2$, MSE, RMSE and $U$ (Thiel's factor), as follows:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y} - \bar{Y})^2}{\sum_{i=1}^{n}(Y - \bar{Y})^2}$$

$$MSE = \frac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{n}}$$

$$U = \frac{\sqrt{\sum_{i=1}^{n}(Y-\hat{Y})^2/n + \sum_{i=1}^{n}(\hat{Y}-\bar{Y})^2/n}}{\sqrt{\sum_{i=1}^{n}(Y-\bar{Y})^2/n}}$$

Where $\hat{Y}$: the predicted value, $\bar{Y}$: the average of the observed data, $Y$: the observed data.

Whenever $R^2$ is between 0 and 1, or higher, the performance is of higher fit [27, 28], while the smallest values of MSE, RMSE and $U$ means a good agreement between detected data and forecast values [28].

4. Results and Dissection

4.1. Elevation – Discharge – Distance Developed Equation for Shatt Al-Hillah

The level of water in shatt Al-Hillah was estimated using Multiple Linear Regression (MLR) model with set of data consisted of 81 observations for Distance (X) and Discharge (Q) of Shatt Al-Hillah. The model was developed using the Statistica10 program. By processing the collected data, two equations were produced which are the following:

$$\hat{E} = 1.949 + 2.562 \times 10 - 4X + 0.299 Q \quad \text{for} \quad E < 28.25 \text{ m.a.s.l. and } Q < 123 \frac{m^3}{s}$$

$$\hat{E} = 1.719 + 9.285 \times 10^{-5}X + 0.140 Q \quad \text{for} \quad E < 28.25 \text{ m.a.s.l. and } Q < 123 \frac{m^3}{s}$$

The limits of above equations are listed in Table 1. The reason for generating two equations to predict the levels of water in Shatt Al – Hillah is the presence of breakpoint at discharge (123 m$^3$/s) which means that errors have maximum values at that point, i.e. if Q value larger than 123 m$^3$/s applied in Equation 15, the predicted values of water levels would be wrong, hence Equation 16 has been developed. The two equations have determination coefficient ($R^2$) equal to 0.932 which means that the observed levels are very close to the fitted line of the case study. In both equations, constants ($ao$) and beta of predicated factors ($\beta_1, \beta_2$) have positive values which means that the increase in discharge and distance will cause increase in levels of water in Shatt Al Hillah.

| Table 1. The limits of level, Distance and Discharge |
|-----------------|---------------------------------|-----------------|-----------------|
|                | Mean                           | Standard division | Min.            | Max.            |
| X (m)           | 30792.91                       | 16027.75          | 0.00000         | 51100.00        |
| Q (m$^3$/s)     | 132.48                         | 56.37             | 27.41500        | 230.00          |
| E (m.a.s.l.)    | 28.25                          | 1.32              | 25.87900        | 31.30           |

The observed data and the corresponding results of predicted levels obtained from Equations 15 and 16 are shown in Figure 2 which gives the best fit line between the observed levels and the predicted ones. The figure shows that Multiple Linear Regression (MLR) model is a very accurate method simulating water level as depending variable (it depends on distance) and discharge as independent variables. The figure also indicates that there is a strong correlation...
between predicted values of water elevation and the real observations.

![Predicted and observed values of elevation (m.a.s.l.) for Shatt Al – Hillah](image)

Figure 2. Predicted and observed values of elevation (m.a.s.l.) for Shatt Al – Hillah

Figure 3 showed the histogram of the residuals, while Figure 4 gave the corresponding normal probability plot of the residuals. The histogram recommends that the residuals are normally distributed, but there are two extreme outliers (larger than 2 and less than – 2). However, the plot demonstrates that the residuals are normally distributed with just two outliers which means that by ignoring these two points, the relationship is approximately linear.

![Histogram of residuals](image)

Figure 3. Histogram of residuals
Correlations between the three variables input in Multiple Linear Regression (E, X and Q) are presented in Figure 5. The figure showed a significant negative correlation between water elevation (E) and distance (X). While, the figure indicated that the significant positive correlation occurred between water elevation (E) and discharge (Q) which make it clear that there is a weak negative correlation between discharge (Q) and distance (X).

Figure 5. The correlation between Elevation (E (m.a.s.l.)), Distance (X (m)) and Discharge (Q (m³/s)) (Using Statistica Program)
Figures 6 (a, b and c) showed the Box and Whisker plots for elevation (m.a.s.l.), distance (m) and discharge (m$^3$/s), respectively. Figure 6a indicated that the elevation values of inter-quartile range from 27.279 m.a.s.l. to 29.139 m.a.s.l. and there are no outliers and the distribution is symmetrically with a slightly positive skew (right skewed distribution). Figure 6b showed that the inter-quartile range for the distance is (23090 m – 46550 m), and there are no outliers too but the distribution have a considerable negative skew (left skewed distribution). Finally, Figure 6c which revealed that the discharge values of the inter-quartile range from 87.415 m$^3$/s to 181 m$^3$/s, and also there are no outliers but the distribution has slightly negative skew (left skewed distribution).
Figure 6(c). Box & Whisker test for discharge (Q (m³/s))

After representing the results obtained from MLR method in the estimation of water elevation of Shatt Al-Hillah, now a rating curve of three gaging stations will be represented and discussed.

4.2. The Rating Curve for the Three Gaging Stations

The construction of the stage – discharge curve (rating curve) for the three primary gauging stations of Shatt Al – Hillah has been produced by applying Simple Linear Regression method (SLR) which represented by Equations 7 to 10. For n = 3372, observations of Elevation (E) and Discharge (Q) for the three primary gauging stations, the following equations are predicted:

- At Hindiya barrage (upstream) at (0.00) kM after calculating α and η and E₀ is given equal to (25 m.a.s.l.), the rating curve was:

  \[
  Q = 0.461 (E - 25.00)^{3.422}
  \]  

  (17)

Figure 7 showed the observed and predicted values of discharge according to the developed rating curve for the whole ten years. The figure indicated that there is a very good agreement between the observed values of water elevation - discharge and the corresponding predicted values i.e. the developed rating curve represents the actual discharge crossing.
that station with the corresponding stage elevation.

- While at the second gauging station (39.00 km from the barrage) (inside Hillah city, \( \alpha \) and \( \eta \) are calculated with Eo given as 23.95 m.a.s.l. the rating curve became as presented in Equation 18 below, and predicted values of discharge with observed ones are shown in Figure 8:

\[
Q = 0.751 (E - 23.95)^{3.474}
\]  

(18)

Figure 8. Computed and observed value of discharge at gauging station inside Hillah City

Figure 8 showed that the calculations of discharges of the first gauging station from the developed Equation 18 are not sophisticated. This can be attributed to the dispersed observed data. Where the flow in the gauging station set up inside Al-Hilla city couldn’t be controlled like in Hindiya Barrage gauging station.

- Finally, at the downstream gage station namely (Dorah regulator station), \( \alpha \) and \( \eta \) are obtained with Eo given as 22.95 m.a.s.l. then the rating curve is:

\[
Q = 3.433 (E - 22.95)^{2.128}
\]  

(19)

The estimated values of discharge from Equation 19 and the observed ones with the measured elevation are showed in Figure 9. The figure showed that there are wide differences between the predicted discharges and the observed ones. This is attributed to the widespread of observed discharges values for a corresponding water elevations in addition to the lack of control on the discharge and water elevation at Dora gauging station at Shatt Al- Hilla River and across 51100 m distance.

Figure 9. Predicted and observed value of discharge at gauging station at (Dora Regulator)
4.3. The Goodness of Fit

From Equations 11 to 14, determination coefficient ($R^2$), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Thiel's factor ($U$) are calculated respectively and represented as best fit. The results are listed in Table 2.

As mentioned earlier, the determination coefficient ($R^2$) ranged from 0 to 1, the closing value of $R^2$ to 1 means that the fitted regression line is much closer to the observed data. All the fitted regression lines developed in this paper represented by Equations 15 to 19 are very satisfactory, where all of them having $R^2$ more than 0.9. Therefore, it is considered that Equations 15 and 16 efficiently represent the stage – distance – discharge relationship. Although each one of rating curves was represented the stage – discharge relationships for the adopted gauging stations. Mean Square Error and Root Mean Square Error represent average squared error and its root for the predictor values. Table 2 illustrated values of both of MSE and RMSE for each developed equations which can be acceptable. These values are 24.095 and 4.908 for stage –distance – discharge relationship and even the values of the rating curves higher than that can be acceptable. This is because of the large number of data used in the regression. In Thiel's factor ($U$), as close as to zero means that the efficiency of the predicted equations are very high. From all the aforementioned information, all the predicted equations were acceptable.

<table>
<thead>
<tr>
<th>Predicted Equation</th>
<th>$R^2$</th>
<th>MSE</th>
<th>RMSE</th>
<th>$U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equations 15 and 16</td>
<td>0.932</td>
<td>24.095</td>
<td>4.908</td>
<td>0.0876</td>
</tr>
<tr>
<td>Equation 17</td>
<td>0.960</td>
<td>562.415</td>
<td>23.715</td>
<td>0.0779</td>
</tr>
<tr>
<td>Equation 18</td>
<td>0.943</td>
<td>748.987</td>
<td>27.367</td>
<td>0.0863</td>
</tr>
<tr>
<td>Equation 19</td>
<td>0.924</td>
<td>925.450</td>
<td>30.426</td>
<td>0.0889</td>
</tr>
</tbody>
</table>

5. Conclusion

The stage – distance – discharge relationship was developed for Shatt Al- Hillah River. This relationship was predicted by using Multiple Linear Regression model solved by Least Square method. This can be used to transform the available data of distance (m) and discharge (m$^3$/s) into stage (water elevation (m.a.s.l.)). Also, the rating curve developed for each one of the three primary gauging stations for Shatt Al- Hillah River using Simple Linear Regression model solved by Least Square method. The method demonstrates a very good agreement between the expected discharge values and the detected ones. So, the advanced rating curves are capable of representing the stage – discharge relationship for each mentioned gauging station.

6. Acknowledgement

The authors would like to acknowledge Babylon Water Resources Department for the help in providing the data. Highly gratitude is given to Prof. Kadhim Naief Kadhim and everyone who has been involved directly or indirectly in assisting and supporting this research.

7. Conflicts of Interest

The authors declare no conflict of interest.

8. References


