Regression-Based Models for Predicting Discharge Coefficient of Triangular Side Orifice

M.K. Elshaarawy¹ and A.K. Hamed² and S.H. Moharram²

¹Demonstrator, Civil Engineering Department, Horus University-Egypt, New Damietta 34517, Egypt
Emails: melshaarawy@horus.edu.eg; akamal@horus.edu.eg
²Professor, Civil Engineering Department, Horus University-Egypt, New Damietta 34517, Egypt
Email: shamed@horus.edu.eg

Abstract - This study introduced another technique to predict the discharge coefficient ($C_d$) of the triangular side orifice (TSO). This technique is based on the SPSS software as multiple linear regression (MLR) and multiple nonlinear regression (MNLR) models. These models were established using 570 experimental datasets, 70 and 30% for calibration and testing stages, respectively. These sets considered five non-dimensional parameters, including (orifice crest height, orifice length, orifice height, upstream flow depth, and Froude number of the main channel). Results showed that the MLR and MNLR models in the calibrating stage had higher determination coefficients and lower errors. In addition, the importance of the input parameters was investigated, showing that the orifice crest height and Froude number highly affect the discharge coefficient value by 36%. In the testing stage, the estimated discharge coefficient by the MLR and MNLR models stayed within the range ±12 and ±5%, respectively, of the experimental values. The MNLR model demonstrated a high level of equivalence compared to previous studies, which provided a mathematical expression to easily predict the TSO's discharge coefficient.

Keywords - Discharge coefficient, Linear regression, Nonlinear regression, Prediction, Side orifice

INTRODUCTION

In open waterways, side weirs, side orifices, and sluice gates redirect some main channel flow to lateral channels and regulate distributary heads. They are also used in irrigation and drainage networks, wastewater treatment systems, aeration basins, and sedimentation tanks [1, 2]. Knowing the diversion structures' water volume requires accurate discharge coefficient estimation. Various experimental and analytical research has estimated flow diversion structures' discharge coefficients. The square side orifice (SSO) was studied by Gill [3] as an example of a particular spatially variable open channel flow. The rectangular side orifice (RSO)'s flow properties were studied by Ramamurthy et al. [4, 5]. They offered a formula that considers the orifice's length, the main channel's width, and the velocity difference between the orifice's jet and the flow.

Masoud [6] studied the water flow patterns of rectangular side sluice gates and discovered that the approach Froude number ($F_r$) impacts the sluice gate discharge coefficient. Hussain et al. [7] conducted experiments on sharp-crested circular side orifices (CSOs) in rectangular main channels. They obtained an equation of 5% accuracy for estimating the discharge coefficient considering the Froude number, orifice diameter, and width of the main channel. Also, Hussain et al. [2] investigated RSO in rectangular main channels and the parameters affecting orifice discharge. The group technique was used to model the flow discharge of RSO by Etehaj et al. [8]. Using the energy equation, Vatankhah and Bijankhan [9] computed a theoretical discharge relation for large and small circular orifices. Also, Vatankhah [10] developed a discharge equation for circular orifices and weirs. Hussain et al. [11] developed a ±5% accurate equation to compute the flow discharge of RSO. Hussain et al. [12] examined flow through CSO flow under free and submerged conditions.

On the other hand, Dutta et al. [13] conducted experiments to better understand the discharge capacity of a sharp-crested circular arc and multi-cycle W-form labyrinth weirs within a rectangular flume under free-flow conditions. They used artificial neural networks (ANNs), support vector machine (SVM), and multiple linear regression (MLR) to create predictive models utilizing the experimental data. They noticed that the SVM model outperformed the others in predicting the discharge accurately. Chakraborty and Goswami [14] predict the factor of safety (FOS) of the slopes using MLR and ANN. Emiroglu et al. [15] built an ANN model to ascertain the triangular labyrinth side weir's discharge coefficient in a straight channel. Their ANN approach outperformed multiple linear and nonlinear regression models in accurately predicting the discharge coefficient.

A few studies have used data-driven strategies to forecast the discharge coefficient of SSOs, CSOs, and RSOs (e.g., [8], [16], [17]). Roushangar et al. [18] explored the combination of SVM with genetic algorithm (SVM-GA) and GEP to estimate the discharge coefficient for trapezoidal and rectangular sharp-crested side weirs. Their findings showed that the SVM-GA model outperformed GEP in terms of accuracy. Eghbalzadeh et al. [16] employed ANNs to estimate the SSOs and CSOs' discharge coefficients. Azimi et al. [19] utilized an adaptive neuro-fuzzy inference system with a genetic algorithm (ANFIS-GA) to predict the RSOs' discharge coefficients. Guo and Stitt [20] developed a technique for determining flow discharge from a partially submerged circular orifice.

Doi: 10.21608 erjeng.2023.244750.1292
Although the different methods adopted by the researchers in their studies, the assessment of the discharge coefficient in the side triangular orifice may need improvement. According to this, the main goal of this study was to use another method based on SPSS software to estimate the discharge coefficient of a side triangular-shaped orifice on a channel. Multiple linear regression (MLR) and multiple nonlinear regression (MNLR) were carried out to obtain the formulas of the discharge coefficient. As a secondary aim, the models' performance was compared with previous studies results to stand the presence of models.

RESEARCH METHODOLOGY

The methodological approach used in the current study can be described as shown in Figure 1. Firstly, dimensional analysis for various combinations of flow and geometric parameters that affect the discharge coefficient \( C_d \) of TSO was investigated. Secondly, 570 experimental datasets were collected to prepare the independent variables for the MLR and MNLR models. Finally, the models were tested to evaluate the performance of these models by comparing the predicted \( C_d \) from the models with the experimental data.

### Dimensional analysis

Equation (1) is a standard equation that can be used to compute the discharge via a free orifice under gravity when the head above the orifice is large compared to the free orifice dimensions [1]. It was given as follows:

\[
Q_s = C_d \frac{LH}{2} \sqrt{2g h_c}
\]

where \( Q_s \) is the discharge through a free orifice \((\text{m}^3 \text{s}^{-1})\), \( g \) is the gravitational acceleration \((\text{m s}^{-2})\), and \( h_c \) is the flow height above the centroid of the orifice section and is calculated in \((\text{m})\) as follows:

\[
h_c = y_c - W - \frac{H}{3}
\]

where \( y_c \) is the flow depth from the water surface to the channel bed at the above side orifice \((\text{m})\), and \( W \) is the orifice crest height \((\text{m})\). Figure 2a provides a 3D view of a TSO placed in a horizontal rectangular channel. Figure 2b illustrates the water surface profile in TSO of length \((L)\) and height \((H)\) under subcritical flow conditions. Vatankhah and Mirnia [1] showed that \( h_1 \) and \( h_2 \) represent the measurements from the free water surface to the TSO crest at the upstream \((U/S)\) and downstream \((D/S)\) ends, while \( y_1 \) and \( y_2 \) represent the corresponding flow depths.

![Figure 2. Triangle side orifice (a) 3D view, and (b) water surface profile view.](image)

### Performance of Models

2.2. Dimensional analysis

The TSO's discharge coefficient \((C_d)\) mainly depends on the flow characteristics, fluid properties, and geometric parameters of TSO. The flow characteristics include upstream flow depth \((y_1)\) and upstream depth-average velocity in the main channel \((V_1)\); fluid properties include water density \((\rho)\), surface tension \((\sigma)\), viscosity \((\mu)\), and gravity acceleration \((g)\). While geometric parameters include orifice height \((H)\), orifice length \((L)\), orifice crest height \((H_c)\), and main channel width \((B)\). Hence, the \( C_d \) of sharp-crested TSO is a function of several parameters, which can be summarized as follows:

\[
C_d = f (\rho, \mu, \sigma, g, L, H, W, B, y_1, V_1)
\]

By applying dimensional analysis using Buckingham theory [21–23], the \( C_d \) of TSO can be a function of the following dimensionless parameters:
\( C_d = f \left( \frac{W B B_1 y V_1}{H L H H' g' y'} \right) \)  

It is possible to express Eq. (4) in the following manner:

\( C_d = f (W^*, L^*, H^*, Y^*, F_r) \)  

2.3. Database description

The experimental data used in the current study to develop MLR and MNLR models were taken from Vatankhah and Mirnia [1]. Overall, 570 experimental datasets were used, with 70\% for training and 30\% for testing. These sets considered combinations of non-dimensional parameters presented in Eq. (5). Table I depicts the descriptive statistics for data collected from Vatankhah and Mirnia [1] experiments.

<table>
<thead>
<tr>
<th>Descriptive statistics</th>
<th>( W^* )</th>
<th>( L^* )</th>
<th>( H^* )</th>
<th>( Y^* )</th>
<th>( F_r )</th>
<th>( C_d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.500</td>
<td>0.625</td>
<td>2.500</td>
<td>0.181</td>
<td>0.178</td>
<td>0.321</td>
</tr>
<tr>
<td>Maximum</td>
<td>2.500</td>
<td>0.833</td>
<td>6.250</td>
<td>0.641</td>
<td>0.824</td>
<td>0.585</td>
</tr>
<tr>
<td>Mean</td>
<td>1.205</td>
<td>0.729</td>
<td>4.039</td>
<td>0.404</td>
<td>0.437</td>
<td>0.500</td>
</tr>
<tr>
<td>Variance</td>
<td>0.397</td>
<td>0.011</td>
<td>2.420</td>
<td>0.013</td>
<td>0.018</td>
<td>0.002</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.630</td>
<td>0.104</td>
<td>1.556</td>
<td>0.115</td>
<td>0.132</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Figure 3 shows the histograms of the collected data. It is obvious that the range discharge coefficient (\( C_d \)) of all datasets is 0.45–0.58. Thus, such models can be developed to predict the \( C_d \) of TSO in the abovementioned range. Figure 3 also demonstrates that the variables used in the database cover a wide range, proving the collected database's reliability. Hence, the proposed models based on such a dataset can accurately predict the coefficient of discharge.
2.4. Regression Models

The Statistical Package for the Social Sciences (SPSS); [24] uses the MLR and MNLR models to develop linear and nonlinear equations, respectively. The equations were calibrated using 70% of the experimental dataset and tested by the remaining 30%.

2.4.1. Multiple Linear Regression

The regression models estimate the level of correlation and determine the dependence among the input and output variables. The MLR models are mostly fitted by the least square approaches. The MLR evaluates the correlation between the response (dependent) from several independent variables and produces a straight-line relationship [25]. The expression of MLR is shown in Eq. (6).

\[
Y = a_0 + \sum_{j=1}^{m} a_j X_j
\]

where \(Y\) is the output of a model \((C_d)\), \(X_j\)'s are the input variables to the algorithms, and \(a_0, a_1, a_2, \ldots, a_m\) are partial regression coefficients.

2.4.2. Multiple Nonlinear Regression

Nonlinear models are simple, interpretable, and predictive [26], [27]. These models can accommodate a wide variety of mean functions. However, they can be less flexible than linear models regarding the data they can describe. However, nonlinear models appropriate for a given application can be more parsimonious (i.e., have fewer parameters) and easier to interpret. Interpretability comes from associating parameters with a biologically meaningful process [28], [29]. MNLR model is applied in the following steps:

1. Defining the dependent variable.
2. Proposing a nonlinear equation in which the dependent variable is a function of the independent variables.
3. Entering the estimation parameters of the proposed nonlinear equation by assuming the starting value; Levenberg-Marquardt was the used estimation method.

4. Finally, the MNLR analysis was started, and the model results were shown in the output log. By trial and error, obtaining the best form of the nonlinear equation developed by the MNLR analysis for the dependent variable \((C_d)\), Eq. (7) is proposed as follows:

\[
C_d = c_1 + c_2 (W^*)^c_3 (L^*)^c_4 (H^*)^c_5 (Y^*)^c_6 (F_r^*)^c_7
\]

where \(c_1, c_2, c_3, c_4, c_5, c_6, \) and \(c_7\) are constant parameters.

2.5. Statistical metrics

The performance of MLR and MNLR models was evaluated by comparing the experimental and predicted discharge coefficients using three statistical metrics (i.e., \(R^2, MAPE, RMSE, NRMSE, \) and \(MAE\)). These metrics can be calculated as follows:

\[
R^2 = \frac{\left(\sum_{i=1}^{n} (\frac{Y_i - \bar{Y}}{\bar{Y}})(\frac{\hat{Y}_i - \bar{Y}}{\bar{Y}})\right)^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2 (\sum_{i=1}^{n} y_i - \bar{y})^2}
\]

\[
MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}
\]

\[
NRMSE = \frac{RMSE}{\bar{Y}}
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|
\]

where \(R^2\) is the determination coefficient, \(MAPE\) is the mean absolute percentage error, \(RMSE\) is the root mean squared error; \(NRMSE\) is the normalized root mean squared error; \(MAE\) is the mean absolute error; \(n\) is the number of a dataset; \(x_i\) is the actual discharge coefficient value; \(\bar{x}_i\) is the average of experimental discharge
coefficients; $y_1$ is the predicted discharge coefficient from the proposed models.

RESULTS AND DISCUSSION

3. Prediction of discharge coefficient via MLR model

The MLR model developed a linear equation and estimated the values of the proposed parameters (i.e., $a_0$, $a_1$, $a_2$, $a_3$, $a_4$, and $a_5$). By applying Eq. (6), the developed linear equation can be written as follows:

$$C_d = 0.881 - 0.074(W^*) - 0.296(L^*) - 0.011(H^*) + 0.025(Y^*) - 0.225(F_r)$$

(13)

3.1. Prediction of discharge coefficient via MNLR model

The MNLR model developed a nonlinear equation and estimated the values of the proposed parameters (i.e., $c_1$, $c_2$, $c_3$, $c_4$, $c_5$, $c_6$, and $c_7$). By substitution in Eq. (7), the developed nonlinear equation can be written as follows:

$$C_d = -0.50(W^*)^{0.350}(L^*)^{0.927}(H^*)^{-0.003}(Y^*)^{0.030}(F_r)^{0.580} + 0.736$$

(14)

3.2. Importance of the independent variables

Figure 4 shows the importance of each independent input variable (i.e., $W^*$, $L^*$, $H^*$, $Y^*$, and $F_r$). It illustrates that the independent variables affected the dependent variable ($C_d$) by 36.13, 26.76, 1.07, 0.06, and 36.98%, respectively. Results showed that the TSO’s discharge coefficient was highly affected by the upstream Froude number ($F_r$) and crest height ratio of TSO ($W^*$), followed by the orifice length ratio ($L^*$). While the orifice height ratio ($H^*$) and orifice height ratio ($Y^*$) were the least important in predicting the discharge coefficient.

3.3. Performance of Models

Table III summarizes the calculated statistical metrics of $R^2$, MAPE, RMSE, NRMSE, and MAE for both models. All statical metrics represent TSO’s experimental and predicted discharge coefficient in good agreement, with higher correlation coefficients and lower error values. Figure 5 also shows the deviation of $C_d$ between the experimental and developed equations in the testing stage. It observed that both MLR and MNLR models stayed within the range of ± 12 and 5%, respectively, from experimental results.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>MAPE</th>
<th>RMSE</th>
<th>NRMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>0.953</td>
<td>0.018</td>
<td>0.011</td>
<td>0.022</td>
<td>0.002</td>
</tr>
<tr>
<td>MNLR</td>
<td>0.996</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Figure 4. Importance of the input variables.
On the other hand, Khosravinia et al. [30] employed three data-driven methods, namely SVM, Least Squares Support Vector Machines (LSSVM) and Gravitational Search Algorithm (LSSVM-GSA). They indicated and concluded that the LSSVM-GSA model had the highest efficacy in predicting the $C_d$ of TSO with $R^2$ and $MAE$ of 0.965 and 0.077, respectively. By comparing the current study with Khosravinia et al. [30], the efficacy of the proposed MNLR was superior, with these models appearing closer to the experimental datasets. Thus, the MNLR is recommended as an effective to estimate and calculate the TSO's discharge coefficient.

CONCLUSIONS

This study utilized MLR and MNLR models to predict the discharge coefficient of the sharp-crested triangle side office (TSO). These models were developed by utilizing 570 datasets from previous experimental study, considering five non-dimensional flow and geometric parameters, 70% of which were used for calibration and 30% for testing. These parameters include orifice crest height, orifice length, orifice height, upstream flow depth, and Froude number of the main channel.

In the calibration stage, MLR and MNLR models had higher $R^2$ values of 0.953 and 0.996, lower $MAPE$ values of 1.8% and 0.2%, and lower RMSE values of 0.011 and 0.001, respectively. While in the testing stage, MLR and MNLR models stayed within the range ±12% and ±5%, respectively, of the experimental values. In addition, the discharge coefficient of TSO was highly affected by upstream Froude number and crest height by approximately 36%. Finally, the developed MNLR model performed very well compared to the laboratory study of Khosravinia et al., 2023, which had higher correlation values and lower errors.

DISCLOSURE STATEMENT

The authors reported no potential conflict of interest.

AUTHOR CONTRIBUTIONS

All authors have equal contributions towards the completion of the research work.

DATA AVAILABILITY STATEMENT

All data underlying the results are available as part of the article, and no additional source data is required.

REFERENCES

[7] A. Hussain, Z. Ahmad, and G. L. Asawa, "Discharge characteristics of sharp-crested circular...


