

GENE EXPRESSION AND MULTIPLE REGRESSION MODELS FOR THE COST OF BRIDGE AND CULVERT

Essam A. Gooda¹ and Mohamed A. Nassar²

¹ Professor of Water resources Engineering, Civil Engineering Dept. Beirut Arab University, Debiyah, Lebanon.
essamgooda@yahoo.com

² Associate professor, Water and Water Structures Eng. Dept., Faculty of Eng., Zagazig University, Zagazig,
nasserzagazig@yahoo.com

ABSTRACT

The present paper is directed to the trouble of the cost comparison between the culvert and the bridge. In addition, a combined effort of cost analysis and modeling approach is presented. The predictions of the cost for both culvert and bridge are presented using Gene Expression Programming (GEP) and Multiple Linear Regression (MLR). Furthermore, The predictions of the cost using (GEP) and (MLR) are compared. The concept of using reference bridge and reference culvert is introduced. GEP approach is able to manage effectively with data gaps. Statistics and scatter plots indicate that the new approach produces acceptably results and can be used as an alternative to the MLR. GEP models predict the cost ratio for all datasets with a relatively higher accuracy (R^2 is 3.0% more than MLR), higher correlation (1% more than MLR), and lower error RMSE (0.007% less than MLR).

Keywords: bridge, culvert, Gene Expression Programming and cost analysis.

1. INTRODUCTION

Prediction of hydraulic variables are extremely required in the design process. Because of difficulty of many phenomena, more strong tools are required for model purposes. When a road crosses a canal or drain, either a bridge or culvert may be a good solution. However, a civil engineer is always seeking for a decision; which one is more economic?. The decision is not an easy task. It depends on several criteria. The factors affecting the choice are many. Yassin (1988), presented a procedure for the economic sizing of box culverts. He formulated a set of 13 accurate dimensionless equations for the estimation of the cost of 13 different box culvert sizes. Bridge was not included in the paper. Mostafa Gooda, (2003), presented two statistical equations to simulate the cost of bridge and culvert. The A review for the comparison between the cost of both the bridge an culvert were presented by Mostafa Gooda, (2003).

Many of the studies were presented to compare between bridge and culvert depending on the multiple linear regression techniques (MLR). Fortunately, relatively new and efficient computational techniques were developed and being applied in most of engineering application. One of these promising techniques is Gene Expression Programming (GEP). GEP was successfully applied in maritime engineering by Kalra and Deo, (2007), Singh et al., (2007), Gaur and Deo, (2008), Ustoorikar and Deo, (2008). The present paper concentrates on the cost comparison of both culvert and bridge. In addition, it aims to model the cost for both structures using Gene Expression Programming (GEP) and compare the results with the multiple regression techniques (MLR). The performance of the present equations are compared to actual cost values. The results allow for the designer engineer to decide which one is more economic, the culvert or the bridge.

2. DESIGN GUIDELINES AND ASSUMPTIONS FOR THE BRIDGE AND CULVERT

The design guidelines have been developed by Mostafa, (2003). The general layouts of a sample of both bridge and culvert are shown in Figs. (1 and 2), respectively. For bridge, the components includes the superstructure (slab, cross girders, and main girders), substructure and wing walls. On the other hand, the culvert's components include the box culvert vent and wing walls. The following assumptions are considered:

- The wing walls are of box type for both bridge and culvert;
- The bearing capacity of soil is: $\sigma = 1.0 \text{ kg/cm}^2$;
- For culvert, only one vent box type is considered;
- The live load considered in design of bridge is 60 ton lorry and a surcharge of 600kg/m^2 is considered in the design of different components of both structures;
- The data of soil properties are: $\gamma_{\text{soil}}=1.8 \text{ t/m}^3$ and $\phi=30^\circ$;
- The compressive strength of reinforced concrete is, $f_{\text{cu}}=325 \text{ Kg/cm}^2$ and the strength of high tensile steel is; $f_s = 1600 \text{ Kg/cm}^2$;
- The Quantity of Portland cement is 400 Kg/m^3 ; and
- The price list for bridge and culvert components is given in table (1).

Table (1) The prices list for the different components of bridge and culvert according to Egyptian market price of 2010 [Mostafa Gooda, E.A., 2003]

Bridge		Culvert	
Component	Price ($\$/\text{m}^3$) $\times 1000$	Component	Price ($\$/\text{m}^3$) $\times 1000$
RC superstructure	326.31	RC walls	252.63
RC walls	252.63	RC floor	200.00

in which: C_b is the bridge cost in (US Dollars); C_{br} is the cost of the reference bridge = 49.6×10^3 \$; H_{br} is the height between road level and bed level of canal or drain for reference bridge case = 4.0 m; S_{br} is the bridge span for reference bridge case = 6.0 m; W_{br} is the road width for reference bridge case = 6.0 m; H is the height between road level and bed level of canal or drain; S is the bridge span; W is the road width; σ_{\square} is the \square soil bearing capacity at site; L_w is the length of wing walls; C_c is the culvert cost in (US Dollars); C_{cr} is the cost of the reference culvert = 49.6×10^3 \$; H_{cr} is the height between road level and bed level of canal or drain for reference culvert case = 4.0 m; W_{cr} is the road width for reference culvert case = 6.0 m; L_{cvr} is the width of clear vent for reference culvert case = 2.0 m; H_{cvr} is the height of clear vent for reference culvert case = 2.0 m; L_{cv} is the width of clear vent; and H_{cv} is the height of clear vent. The dimensionless relationships of the bridge and culvert costs and other independent parameters could be determined relative to the characteristics of the reference unites as following:

$$C_{bratio} = f(H_{bratio}, S_{bratio}, W_{bratio}) \dots\dots\dots(3)$$

$$C_{cratio} = f(H_{cratio}, W_{cratio}, L_{cvratio}, H_{cvratio}) \dots\dots\dots(4)$$

In which: $C_{bratio} = C_b/C_{br}$ is the relative bridge cost; $H_{bratio} = H/H_{br}$ is the relative height between road level and bed level of canal or drain; $S_{bratio} = S/S_{br}$ is the relative bridge span; $W_{bratio} = W/W_{br}$ is the relative road width; $C_{cratio} = C_c/C_{cr}$ is the relative culvert cost; $H_{cratio} = H/H_{cr}$ is the relative height between road level and bed level of canal or drain; $W_{cratio} = W/W_{cr}$ is the relative road width; $L_{cvratio} = L_{cv}/L_{cvr}$ is the relative width of clear vent; $H_{cvratio} = H_{cv}/H_{cvr}$ is the relative height of clear vent.

4. PROPOSED DESIGN SOFTWARE

Two different computer programs are designed, well verified and calibrated to perform several procedures for the design of different components of both bridge and culvert, Mostafa, E.A. 2003. The outputs of the programs give:

- Straining actions;
- Stress distribution;
- Stability calculation against sliding and overturning;
- Complete structural design for different components; and
- Cost of the unit. (i.e., the bridge or culvert)

For bridges, several values of H , the difference between the road level and bed levels, are considered. These values are, $H = 4, 4.5, 5, 5.5, 6, 6.5$ m. For each value of H , all bridge

components are estimated and consequently, the cost is determined based on the price list shown in Table (1). Also, different road widths, $W = 6, 8, 10$ m, and different span values, $S = 6, 8, 10$ m, are considered. On the other hand for culverts, same height values; $H = 4, 4.5, 5, 5.5, 6, 6.5$ m and same road widths, $W = 6, 8, 10$ m are considered. Moreover, several dimensions of the clear vent width, L_{cv} , and the height of clear vent, H_{cv} are used, (i.e., $L_{cv} = 3$; and 2; $H_{cv} = 3$; and 2). Characteristics of bridges and culverts as well as their estimated costs are summarized for 15-selected cases in Table (2). Costs of 54 different bridges and costs of 72 different culverts are used in this paper.

5	1.33	1.00	1.50	2.33	1.67	1.13	1.00	1.00	1.10
---	------	------	------	------	------	------	------	------	------

5. MULTIPLE LINEAR REGRESSION (MLR)

Based on the design data, several statistical equations were used to predict the relative bridge cost; C_{bratio} and the relative culvert cost; C_{cratio} . The following equations represent non-dimensional form of the bridge and culvert cost ratio; respectively.

$$C_{bratio} = H_{bratio}^{1.793} \times S_{bratio}^{0.354} \times W_{bratio}^{0.472} \dots\dots\dots(5)$$

$$C_{cratio} = H_{cratio}^{0.289} \times W_{cratio}^{0.201} \times L_{cratio}^{0.180} \times H_{cvratio}^{0.256} \dots\dots\dots(6)$$

Equation (5) represents a non-dimensional form can be applied to estimate the cost ratio of an existing bridge relative to that of the reference bridge. On the other hand, equation (6) represents a non-dimensional form can be applied to estimate the cost ratio of an existing culvert relative to that of the reference culvert. The objective function (i.e., the function is used to measure the accuracy of the MLR equations) values are calculated from the proportion of variance, which is the best single measure of how well the predicted values match the actual values. This is also known as the "coefficient of determination RSQ or R²" as presented in Eq. (7). In addition, the correlation factor is used as a second measure factor. The quantity r , called the *linear correlation coefficient*, measures the strength and the direction of a linear relationship between two variables. The mathematical formula for computing r can be presented as shown in Eq. (8). The last measure to the accuracy of the equations (5) and (6) is the Root Mean Square Error (i.e., RMSE). It can be calculated as presented in Eq. (9).

$$R^2 = \frac{\sum_i^j (Y - \bar{X})^2}{\sum_i^j (X - \bar{X})^2} \dots\dots\dots(7)$$

$$Correlation (r) = \frac{N \sum XY - (\sum X \sum Y)}{((N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2))^{0.5}} \dots\dots\dots(8)$$

$$RMSE = \sqrt{\frac{\sum_i^j (Y - X)^2}{N}} \dots\dots\dots(9)$$

where Y is the predicted variable, X is the measured variable, \bar{X} is the sample mean of X value and N the total number of variables. Figures (3A and 3B) show a comparison between the calculated C_{bratio} and C_{cratio} and the predicted values using Eqs. (5) and (6), respectively. A good agreement can be noticed, to some extent. The residuals of the previous equations are plotted versus the predicted values as shown in (3a and 3b). The residuals show random distribution around the line of zero in case of equation (5), while it follows a specified trend

for equation (6). The regression statistics have been listed in the table (3). Results of MLR emphasize the need to adopt a new technique like Gene Expression Programming (GEP).

Table (3) Basic Features of the developed MLR models equations (5) and (6).

MLR Models	R ²	RMSE	Correlation factor
Equation (5)	0.941	0.092	0.990
Equation (6)	0.878	0.034	0.938

Dividing equation (5) by equation (6) gives the following equation:

$$\frac{C_{bratio}}{C_{cratio}} = \frac{H_{bratio}^{1.793} \times S_{bratio}^{0.354} \times W_{bratio}^{0.472}}{H_{cratio}^{0.2893} \times W_{cratio}^{0.2013} \times L_{cvratio}^{0.1801} \times H_{cvratio}^{0.2560}} \dots\dots\dots(7)$$

As mentioned before, the costs of both reference bridge and reference culvert are the same. Therefore: $C_{bratio}/C_{cratio} = C_b/C_c$. This means that:

$$\frac{C_b}{C_c} = \frac{H_{bratio}^{1.793} \times S_{bratio}^{0.354} \times W_{bratio}^{0.472}}{H_{cratio}^{0.2893} \times W_{cratio}^{0.2013} \times L_{cvratio}^{0.1801} \times H_{cvratio}^{0.2560}} \dots\dots\dots(8)$$

For the same road width, and the same height between road level and bed level of canal or drain, in both bridge and culvert, equation (8) may take the following form:

$$\frac{C_b}{C_c} = \frac{H^{1.503} \times W^{0.270} \times S_{bratio}^{0.354}}{L_{cvratio}^{0.1801} \times H_{cvratio}^{0.2560}} \dots\dots\dots(9)$$

In which: C_b is the bridge cost in (US Dollars); C_c is the culvert cost in (US Dollars); H is the height between road level and bed level of canal or drain; and W is the road width; S_{bratio} is the relative bridge span; $L_{cvratio}$ is the relative width of the culvert clear vent; $H_{cvratio}$ is the relative height of the culvert clear vent..

6. ANALYSIS OF THE OUTPUTS USING THE MULTIPLE LINEAR REGRESSION (MLR)

The effect of the height between road level and bed level of canal or drain on the cost for both bridge and culvert is shown Figs. (5, 6 and 7) for $W = 6, 8, 10$ m, respectively. For bridge, the rate of change is steeper than that for culvert. The trend is logic where increasing the value of H or bridge height, leads to increase height and thickness of not only abutment but also wing walls. For culvert, H is just a fill height above culvert. In contrast, it can be noticed that there is a short range for H at which the cost of the bridge may be more economic in comparison with the culvert cost. The results of regression analysis show that for bridge, the power, of H is 1.793 while it is 0.289 in case of culvert. The effect of road

width, W is shown in figure (8). For example, the “bridge to culvert cost ratio” changes from 1.0 at $W = 6$ m to 1.2 at $W = 10$ m, in case of $H = 4$ m and $S = 6$ m.

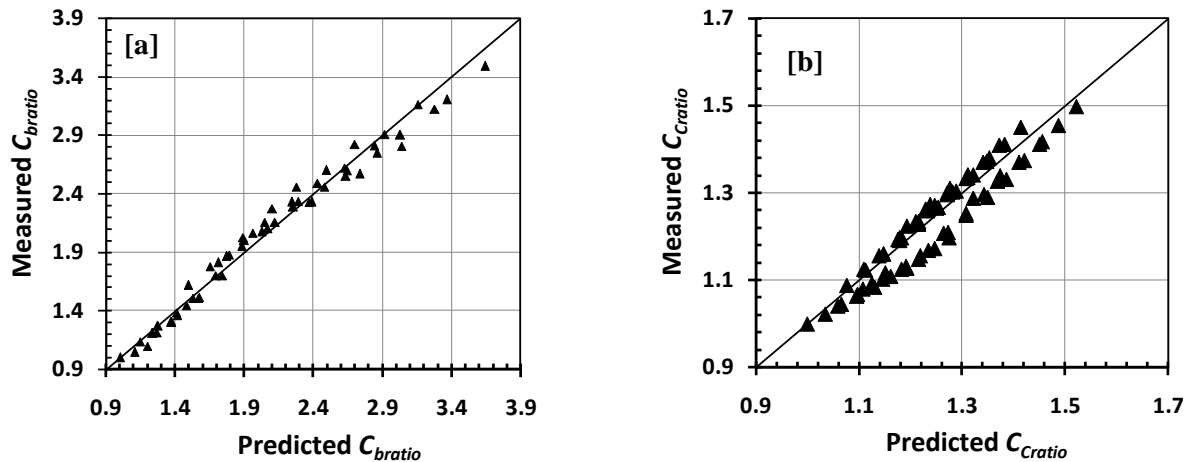


Fig. 3 Comparison between the different calculated cost ratios and the predicted ones using Eqs. (5) and (6), respectively: [a] C_{bratio} [b] C_{cratio}

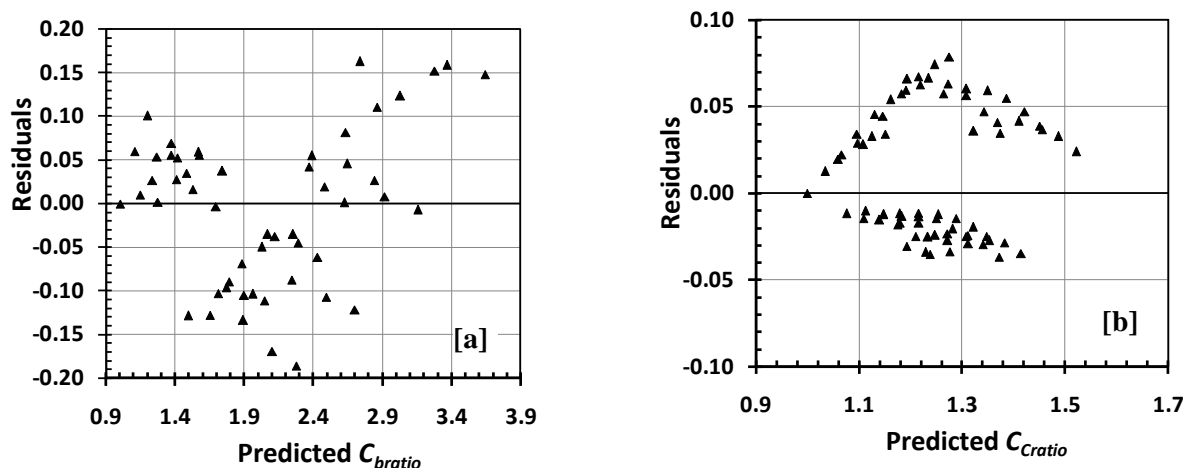


Fig. 4 Variations of residuals for different data sets of the predicted cost ratios using Eqs. (5) and (6), respectively: [a] C_{bratio} [b] C_{cratio}

7. GENE EXPRESSION PROGRAMMING (GEP) MODELS

GEP technique, which is an extension to genetic programming (GP), used programs all encoded in linear chromosomes, which are then expressed into Expression Trees (ETs). ETs are difficult computer programs that are usually used to simulate a particular case and are chosen according to their fitness for solving that case. Once the guidance set was chosen, one could say that the knowledge location of the system is defined. A population of candidate programs is formed and then each program is tested against a predefined fault. The independent parameters, C_{bratio} , and C_{cratio} were selected as the output terminal. To find the optimum formulation, eight functions, namely, plus, minus, multiplication, division, negative, inverse, log and sin were used.

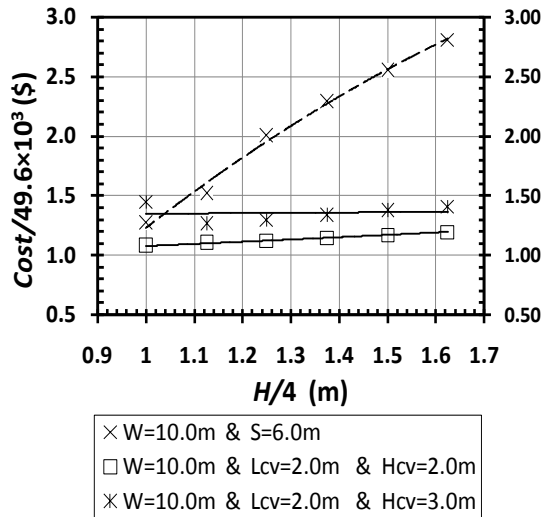


Fig. 7 The effect of H on the cost for both bridge and culvert for $W=10m$

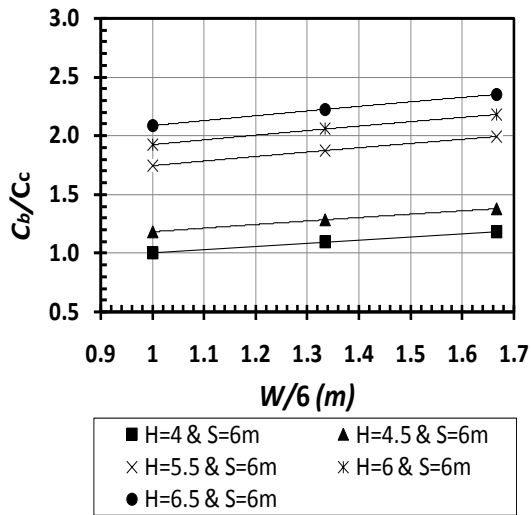


Fig. 8 The effect of W on the bridge to culvert cost ratio for $S=6m$ and different H

A large number of generations were needed to find a principle with a minimal fault. First, the some constants were assigned. With these constants, a large number of generations were required to reduce the fault. These constants were changed and the program was executed to search for a principle or formula with minimal fault and as short as possible in length. The optimum GEP structure had the following characteristics.

- [1] The selection is done by taken a random number of individuals from the population and the best fit is chosen;
- [2] The operations that were used in this study were crossover and mutation. They were selected by adopting a rule with a minimum probability of 0.1;
- [3] The sum of absolute differences between the obtained and expected values for all sets of data in the database was used as a measure for fitness;
- [4] Population size: 50 members;
- [5] Maximum Gene head length 8.0, [6] gene per chromosome 4.0 and, [6] Total

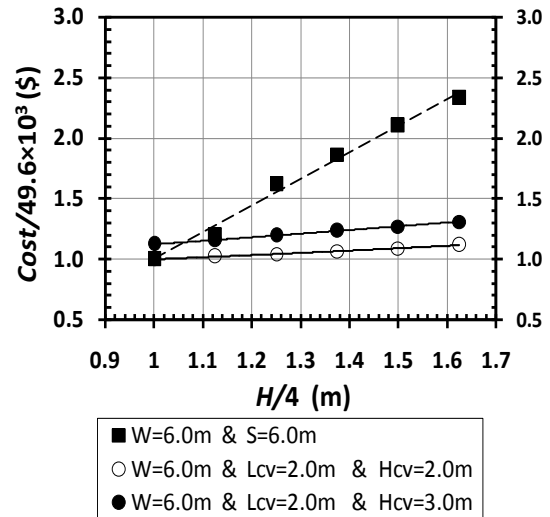


Fig. 5 The effect of H on the cost for both bridge and culvert for $W=6m$

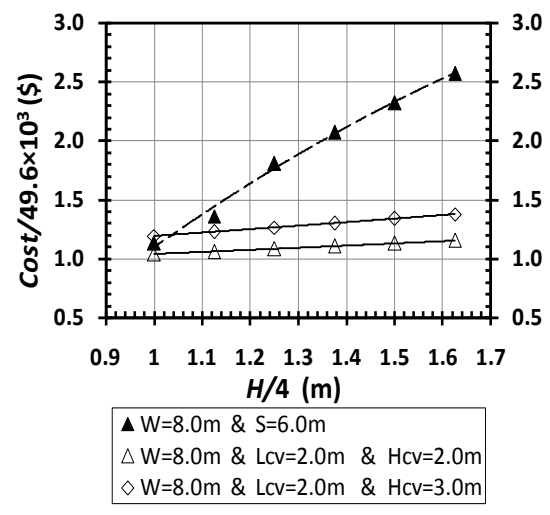


Fig. 6 The effect of H on the cost for both bridge and culvert for $W=8m$

generations are 2000. After incorporating the corresponding values and making necessary simplifications, the final equations become:

$$C_{bratio} = \frac{W_{bratio} - S_{bratio}}{W_{bratio}} + H_{bratio} - \sin W_{bratio} + S_{bratio} H_{bratio} \dots\dots\dots(10)$$

$$C_{cratio} = \log((L_{cvratio} W_{cratio})^{0.25}) + (H_{cvratio} H_{cratio})^{0.25} \dots\dots\dots(11)$$

With the all data sets used in this study, approximately 100 percent of these patterns chosen till the best training performance were seen. Table 4, presents the compiled measurements. They are graphically shown in Fig. (9) which shows the ordinates of the calculated independent parameters against the predicted ones. Presence of a small scatter between these variables can be noted. The statistic measures for both equations (10 and 11) have been listed in the table (4).

Table (4) Basic Features of the developed GEP models of equations (10 and 11)

GEP Models	R ²	RMSE	Correlation factor
Equation (10)	0.967	0.012	0.992
Equation (11)	0.885	0.033	0.944

8. ANALYSIS OF RESULTS

The independent parameters predictions in the present study has been made on the basis of outputs data made earlier in Mostafa, E.A. 2003. The Genetic Expression Programming (GEP), and Multiple Linear Regression (MLR) models, therefore, developed with the former sets of values as inputs (variables namely, H_{bratio} , S_{bratio} , W_{bratio} in order to predict C_{bratio} and C_{cratio} , the cost ratio for both bridge and culvert, respectively.

The statistical results of models predictions for data sets are given in Tables (3, and 4). The results of the developed GEP prediction models were compared with the regression equation formulae (5 6). It was found that the Genetic Expression Programming (GEP) models are highly satisfactory as seen in Figs. (11, 12,13,14,15 and 16). From the comparison of the different predictions models, it is clear that GEP models predicted the cost ratio for all datasets with a relatively higher accuracy (R^2 is 3.0% more than MLR), higher correlation (1% more than MLR), and lower error RMSE (0.007% less than MLR). The acceptable results of GEP are achieved for simulating the other studied parameters compared to the MLR models. Further, the scatter plot Fig. (10) proves out the good performing of the GEP.

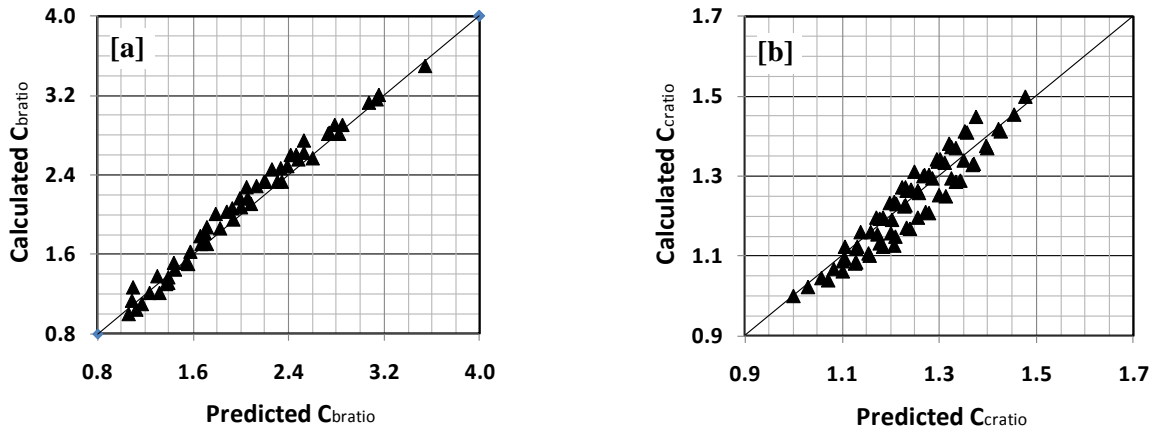


Fig. 9 Comparison between the different calculated cost ratios and the predicted ones using Eqs. (10 and 11), respectively: [A] C_{bratio} [B] C_{cratio}

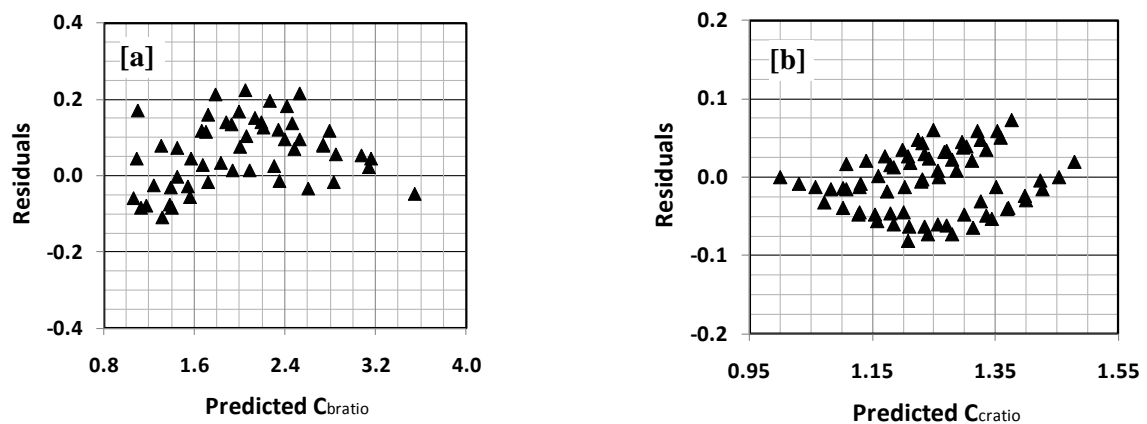


Fig. 10 Variations of residuals for different data sets of the predicted cost ratios using Eqs. (10 and 11), respectively: [A] C_{bratio} [B] C_{cratio}

9. APPLIED EXAMPLES

9.1. Example 1

At certain location, estimate the cost of construction for a bridge has the following characteristics using MLR equation (5) and GEP equation (10):

- Road width, $W=10$ m,
- Span, $S=10$ m, and
- Road height above bed, $H=6$ m

9.2. Solution

- The first step is to estimate the relative values:
 - $W_{bratio}=10/6=1.67$, $S_{bratio}=10/6=1.67$, and $H_{bratio}=6/4=1.5$
- Applying MLR equation (5) to estimate the relative cost of bridge

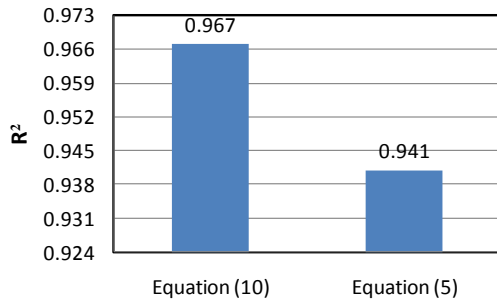


Fig. 14 Comparison between GEP Eq. 10 and MLR Eq. 5 for R²

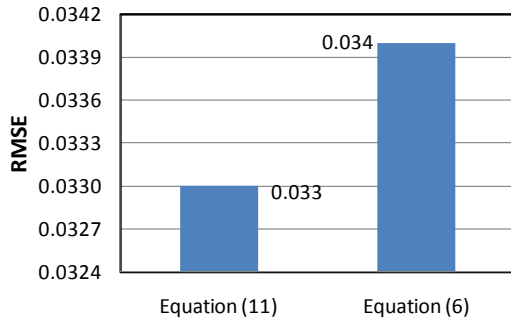


Fig. 15 Comparison between GEP Eq. 11 and MLR Eq. 6 for RMSE

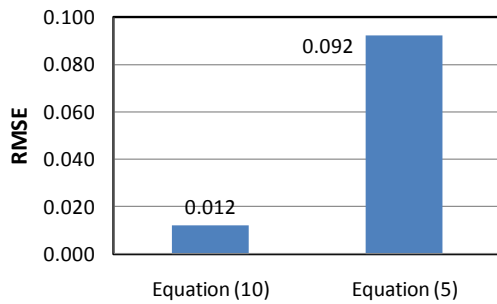


Fig. 16 Comparison between GEP Eq. 10 and MLR Eq. 5 for RMSE

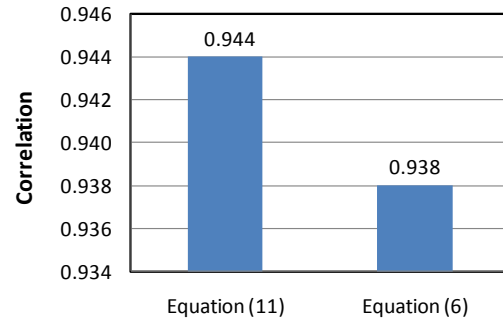


Fig. 11 Comparison between GEP Eq. 11 and MLR Eq. 6 for the correlation coefficient

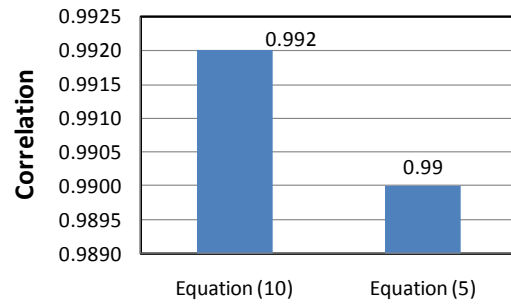


Fig. 12 Comparison between GEP Eq. 10 and MLR Eq. 5 for the correlation coefficient

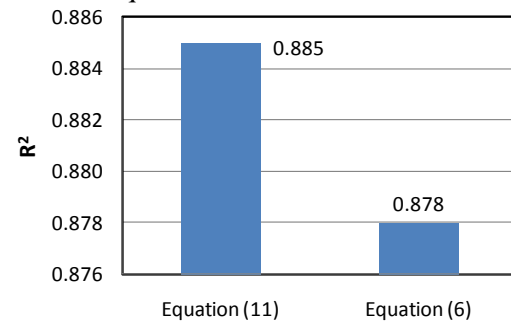


Fig. 13 Comparison between GEP Eq. 11 and MLR Eq. 6 for R²

$$\circ C_{bratio} = H_{bratio}^{1.793} \times S_{bratio}^{0.354} \times W_{bratio}^{0.472} = (1.5)^{1.793} \times (1.67)^{0.354} \times (1.67)^{0.472}$$

$$C_{bratio} = 3.157$$

$$\circ \text{From the actual data this value is } = 300.29/94.95 = 3.16$$

- Applying GEP equation (10) to estimate the relative cost of bridge

$$\circ C_{bratio} = \frac{1.67 - 1.67}{1.67} + 1.5 - \sin(1.67) + (1.67 \times 1.5)$$

$$\circ C_{bratio} = 3.97$$

- These values mean that the existing bridge costs equal 3.157 and 3.97 times that of the reference bridge using MLR equation (5) and GEP equation (10), respectively . It

is known that, $C_{bratio} = C_b/C_{br}$

- C_{br} is the cost of the reference bridge = 49.6×10^3 \$;
- Then, the bridge cost; $C_b = 156.58 \times 10^3$ \$ and 196.9×10^3 \$ using MLR equation (5) and GEP equation (10), respectively.

9.3. Example 2

The above data of bridge given in Example 1 is to be compared with a culvert of the following characteristics:

- Road width, $W=10$ m,
- Road height, $H=6$ m,
- Width of internal vent, $L_{cv}=3$,
- Height of internal vent, $H_{cv}=3$ m,

9.4. Solution

- In example 1, the cost of bridge is estimated. The same procedures can be repeated to estimate the cost of culvert as following;
- $H_{cratio} = 1.5; W_{cratio} = 1.67; L_{cvratio} = 1.5$ and $H_{cvratio} = 1.5$
- Applying MLR equation (6) to estimate the relative cost of culvert
 - $C_{cratio} = 1.5^{0.289} \times 1.67^{0.201} \times 1.5^{0.180} \times 1.5^{0.256} = 1.487$
 - From the actual data this value is $= 139.46/959 = 1.45$
- Applying GEP equation (11) to estimate the relative cost of culvert
 - $C_{cratio} = \log((1.5 \times 1.67)^{0.25}) + (1.5 \times 1.5)^{0.25} = 1.324$
- These values mean that the existing culvert costs equal 1.487 and 1.324 times that of the reference culvert using MLR equation (6) and GEP equation (11), respectively. It is known that, $C_{cratio} = C_c/C_{cr}$
- C_{cr} is the cost of the reference bridge = 49.6×10^3 \$;
- Then, the bridge cost; $C_c = 73.75 \times 10^3$ \$ and 65.69×10^3 \$ using MLR equation (6) and GEP equation (11), respectively.

10. CONCLUSION

The present paper is directed to the trouble of the cost comparison between the culvert and the bridge. The predictions of the cost for both culvert and bridge are modeled using Multiple Linear Regression (MLR) and Gene Expression Programming (GEP). A non-dimensional relationships are presented to estimate the different cost ratios for both bridge

and culvert. The results of the developed MLR show that the relationships are simple and easy to predict the effect of secondary factor on the main factor. GEP prediction models are compared with the regression equation formulae. It was found that the genetic expression programming models are highly satisfactory. From the comparison of the different predictions models, it is clear that GEP models predicted the cost ratio for all datasets with a relatively higher accuracy, higher correlation and lower error RMSE. The acceptable results of GEP are achieved for simulating the other studied parameters compared to the MLR models. Two examples are presented to explain the procedures of estimating the cost of either a bridge or a culvert.

REFERENCES

- [1]. “Hydraulic Charts for the selection of Highway Culverts,” Hydraulic Engineering Circular No. 5, U.S.A. Federal Highway Administration, Washington, D.C., 1965.
- [2]. Capacity Charts for the Hydraulic Design of Highway Culverts Hydraulic Engineering Circular No. 10, U.S.A. Federal Highway Administration, Washington, D.C., 1972.
- [3]. Carter, R.W., Computations of Peak discharge at Culverts,” U.S. Geological Survey, Circular 376, 1957.
- [4]. Chow, yen Te, open Channel Hydraulics,” McGraw—Hill Book Company, 1959.
- [5]. Wright, P.H. and Paquette, R.J., “Highway Engineering,” John Wiley & Sons, Fourth Edition, 1979.
- [6]. Yassin, A. A., “Economical Design of Box Culverts,” Alexandria Engineering Journal, Vol. 27 No. 4, pp. 1-20, October 1988.
- [7]. Young, G.K., Childrey, M.R. and Trent, R.E., “Optimal Design For Highway drainage Culverts” Journal of the Hydraulics Division, ASCE, Vol. 100, No. HY5, 1974.
- [8]. Kalra, R., Deo, M.C., 2007. “Genetic Programming to retrieve missing information in wave records along the west coast of India”. Applied Ocean Research, Vol. 29, No. 3, pp.99–111.
- [9]. Singh, A.K., Deo, M.C., SanilKumar, V., 2007. “Neural network— genetic programming for sediment transport”. Journal of Maritime Engineering, The Institution of Civil Engineers, London, Vol. 160, (MA3), pp.113–119.
- [10]. Gaur, S., Deo, M.C., 2008. “Real time wave forecasting using genetic programming”. Journal of Ocean Engineering, Vol. 35, pp.1166–1172.
- [11]. Ustoorikar, K., Deo, M.C., 2008. “Filling up gaps in wave data with genetic”. Journal of Programming Marine Structures, Vol. 21, pp.177–195.

- [12].Gooda, E.A., 2003 “Cost Analysis for Bridge and Culvert”, Seventh International Egyptian Water Technology Conference, CAIRO, EGYPT, JUNE 3-5, 2003.