Application of a soft computing technique in predicting the percentage of shear force carried by walls in a rectangular channel with non-homogeneous roughness

Zohreh Sheikh Khozani, Hossein Bonakdari and Amir Hossein Zaji

ABSTRACT

Two new soft computing models, namely genetic programming (GP) and genetic artificial algorithm (GAA) neural network (a combination of modified genetic algorithm and artificial neural network methods) were developed in order to predict the percentage of shear force in a rectangular channel with non-homogeneous roughness. The ability of these methods to estimate the percentage of shear force was investigated. Moreover, the independent parameters’ effectiveness in predicting the percentage of shear force was determined using sensitivity analysis. According to the results, the GP model demonstrated superior performance to the GAA model. A comparison was also made between the GP program determined as the best model and five equations obtained in prior research. The GP model with the lowest error values (root mean square error (RMSE) of 0.0515) had the best function compared with the other equations presented for rough and smooth channels as well as smooth ducts. The equation proposed for rectangular channels with rough boundaries (RMSE of 0.0642) outperformed the prior equations for smooth boundaries.

Key words | genetic programming, open channel, roughness, shear

INTRODUCTION

Boundary shear stress is an important parameter in steady, fully developed open channel flow. Trustworthy prediction of boundary shear distribution in open channel flow is very difficult in many critical engineering problems, such as sedimentation, channel design and energy loss calculation. The boundary shear stress distribution in rectangular channels with smooth or rough boundaries has been investigated by many researchers, such as Knight (1981), Knight et al. (1984) and Bilgil (2005). Due to difficulties in measuring shear stress distribution in channels by direct and indirect methods, a number of studies have been extended in order to calculate this by different approaches. Some studies have focused on numerical or analytical means of predicting shear stress distribution (Berlamont et al. 2003; Guo & Julien 2005; Yu & Tan 2007; Ansari et al. 2011; Bonakdari et al. 2015; Sheikh & Bonakdari 2015).

Nowadays, soft computing methods of forecasting various phenomena in all fields are widely exploited. Alternatively, soft computing techniques, such as artificial neural networks (ANNs), evolutionary computation, fuzzy logic, genetic programming (GP) and gene-expression programming, have been successfully applied for water engineering problems in recent years (Azamathulla & Jarrett 2013; Ehtehaj & Bonakdari 2014; Zaji & Bonakdari 2015; Azamathulla 2015). Boundary shear force carried by walls is a major parameter related to mean shear stress. Only a few studies have addressed predicting shear stress in channels using soft computing methods. Cobaner et al. (2010) investigated the ability of the ANN approach to predict the percentage of shear force acting on walls (%SFw) in smooth open channels and ducts. The authors indicated that the ANN model’s function was better than previously obtained equations of other researchers. In this study, the capability of GP and genetic artificial algorithm (GAA) models in predicting the %SFw in rectangular channels with non-homogeneous roughness is investigated. Consequently, two models are developed and the best one is selected. The function of the best model is also compared with equations obtained by Knight (1981), Knight et al. (1984, 1994) and Seckin et al. (2006) for rectangular channels, and an equation proposed by Rhodes & Knight (1994) for smooth ducts.
MODELS

GP model

GP was introduced by Koza (1994) and is among the more practical genetic algorithm (GA) applications. The GP method is very similar to the GA, except that the chromosomes are in fact computer programs. GP begins with a random initial population consisting of certain computer programs. Each computer program is run to evaluate the cost of each chromosome. The cost function is calculated using fitness functions. Thus, by sorting out the initial population costs, and performing crossover, mutation and elite GA processes, a new population is achieved. The main objective of GP is to find a computer program that can accurately predict shear force using the given input variables.

Three major variables affect GP performance: (i) selecting the appropriate input variables; (ii) selecting the functions that GP is permitted to use in the computer programs (functions are arithmetic operators such as + and – or mathematical functions such as exp, sin, power and logical functions like OR and NOT; in this study, four different function combinations are tested); and (iii) selecting a suitable fitness function (a fitness function is used to evaluate the efficiency of each individual, and selecting the appropriate fitness function directly affects the model’s performance). The mean squared error (MSE) and mean absolute error (MAE) statistics methods are herein regarded as fitness functions.

GAA model

The ANN is as a renowned soft computing method employed in diverse engineering problems. A typical ANN is formed of three or four layers: one input layer, one or two hidden layers and one output layer. Each layer consists of neurons, each of which counts the weighted summation of the neurons in the previous layer. After placing this summation into the transfer function, each neuron conveys the results to the neurons in the next layer. The weight of each neuron is determined in the training process. In this study, the Levenberg–Marquardt (LM) algorithm (Levenberg 1944) is used to train the ANN model. The sigmoid transfer function has been successfully used in various ANN simulations; two commonly used types are the logarithmic transfer function (Equation (1)) and the hyperbolic tangent transfer function (Equation (2)). In the present study, these functions are examined for the hidden layers, and the linear transfer function (Equation (3)) is used for the output layer.

\[
\text{logsig}(x) = \frac{1}{1 + e^{-x}} \quad (1)
\]

\[
\text{tansig}(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (2)
\]

\[
\text{purelin}(x) = x \quad (3)
\]

The number of neurons in both input and output layers is the same as the number of input and output variables. One of the most problematic stages in ANN modeling is determining the number of nodes in the hidden layer. Thus, in the GAA method, the GA is used to optimize the ANN model structure. In addition, due to the random nature of the LM training algorithm, ANN models with the same training conditions may produce different results. Therefore, the GA algorithm is modified such that the elite population is repeatedly run in order to determine the most appropriate performance of each chromosome. The performance of each chromosome is calculated using fitness functions. To find a suitable fitness function, the two methods considered are MSE and MAE.

DATA USED

Knight (1981) studied boundary shear stress in a rough rectangular channel by using the Preston tube technique and measured the shear stress of walls and beds with different roughness grades and at different depths. The percentage of shear force carried by walls and beds was also calculated in Knight’s study. Knight (1981) conducted experiments in a flume 15 m long, 460 mm wide and with a constant bed slope of 9.58 × 10^{-4}. The author assumed that \(\%SF_w\) varies exponentially with aspect ratio, \(B/h\), and relative roughness, \(k_{sb}/k_{sw}\), as follows:

\[
\%SF_w = e^\alpha \left(\tanh(\pi \beta) - 0.5(\tanh(\pi \beta) - \beta)^2\right) \quad (4)
\]

where \(\alpha\) is a function of aspect ratio:

\[
\alpha = -3.264 \log \left(\frac{B}{h} + 3\right) + 6.211
\]

\[
\beta = 1 - \frac{\gamma}{5} \quad \text{and} \quad \gamma = \log \left(\frac{k_{sb}}{k_{sw}}\right).\]

Knight et al. (1984) analyzed their own data as well as that of Knight (1981), plotted them on a log-log scale, assumed a
simple relationship between $\%SF_w$ and $B/h$, and derived the following equation:

$$\log(\%SF_w) = A_1 \log\left(\frac{B}{h} + A_2\right) + A_3$$  (5)

where $A_1 = -1.4026$, $A_2 = 3.00$ and $A_3 = 2.6692$.

Equation (5) can be rewritten in another form as:

$$\%SF_w = e^a$$  (6)

where $a = -3.230 \log\left(\frac{B}{h} + 3\right) + 6.146$.

Knight et al. (1994) suggested a general equation for smooth rectangular channels considering the relationship between $\%SF_w$ and the wetted perimeter ratio, $P_b/P_w$, as follows:

$$\%SF_w = \exp\left(-3.23 \log\left(\frac{P_b}{1.38P_w}\right) + 4.6052\right)$$ for $\frac{P_b}{P_w} < 4.374$  (7)

$$\%SF_w = (0.6603(\frac{P_b}{P_w})^{0.28125}) \exp\left(-3.23 \log\left(\frac{P_b}{1.38P_w}\right) + 4.6052\right)$$ for $\frac{P_b}{P_w} \geq 4.374$  (8)

Seckin et al. (2006) experimentally studied boundary shear stress and shear force distributions in smooth rectangular channels. They derived a nonlinear regression-based equation from the experimental analysis to obtain the percentage of total shear force carried by walls as follows:

$$\%SF_w = \exp\left(-3.183 \log\left(\frac{B}{h} + 5\right) + 6.1175\right)$$  (9)

For smooth rectangular ducts, Rhodes & Knight (1994) carried out a series of experiments considering aspect ratios of up to 50 and proposed the following equation:

$$\%SF_w = \frac{100}{1 + \left(1 + 1.345\left(h/B\right)\right)^{-1.097}}$$  (10)

According to these relations, some parameters, such as channel geometry ($B$), flow depth ($h$), bed and wall roughness ($k_{sb}$, $k_{sw}$), energy slope ($S_l$), flow velocity ($V$), fluid density ($\rho$), gravitational acceleration ($g$) and hydraulic radius ($R$) affect the shear force carried by walls. Thus, $\%SF_w$ can be expressed as:

$$\%SF_w = F(\rho, g, B, h, k_{sb}, k_{sw}, R, S_l, V)$$  (11)

By using Buckingham’s theorem, the dimensionless parameters affecting $\%SF_w$ are represented as follows:

$$\%SF_w = f\left(\frac{B}{h}, Fr, \frac{k_{sb}}{k_{sw}}, Re\right)$$  (12)

where Fr is the Froude number, and Re the Reynolds number. About 75% of all data were selected randomly for training and the remaining were used for testing.

### ANALYSIS OF GP AND GAA MODELS

Since the data were distributed over different ranges, all data were initially normalized to zero mean and units with the following equation:

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}}$$  (13)

where $X_n$ is the normalized value, and $X_{min}$ and $X_{max}$ are the minimum and maximum values of the variables, respectively. Subsequently, in order to select the best among the GP and GAA models, different input combinations were investigated. Eight different combinations were grouped as quaternary, ternary and binary, as follows: (1) $B/h$, $Fr$, $Re$ and $k_{sb}/k_{sw}$, (2) $B/h$, $Fr$ and $Re$, (3) $B/h$, $Fr$ and $k_{sb}/k_{sw}$, (4) $B/h$, $Re$ and $k_{sb}/k_{sw}$, (5) $Fr$, $k_{sb}/k_{sw}$ and $Re$, (6) $B/h$ and $k_{sb}/k_{sw}$, (7) $B/h$ and $Fr$ and (8) $Fr$ and $k_{sb}/k_{sw}$. The root mean square error (RMSE) served to check the models’ accuracy in each step. The RMSE, MSE, and MAE statistical indexes are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_p - X_m)^2}{n}}$$  (14)

$$MSE = \frac{\sum_{i=1}^{n} (X_p - X_m)^2}{n}$$  (15)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |X_p - X_m|$$  (16)

where $X_p$ is the wall shear stress predicted by the model, $X_m$ is the wall shear stress measured in the laboratory and $n$ is the number of studied dataset samples.

Figure 1 presents the RMSE values for each input combination in the GP and GAA models. According to this figure, the GP model has smaller input combination values than the GAA model. Moreover, the GP with input combination (2) and RMSE of 0.0530 and GAA with input combination (5).
Upon selecting the best input combination for each model, the best fitness function was selected. In the GP model, input combination (2) and the default function \((+,-,\times,\div)\) were considered and two fitness functions MSE and MAE were tested. The results indicated that for the GP model, the MAE fitness function with RMSE of 0.0521 outperformed the MSE fitness function with RMSE of 0.053. However, the MSE fitness function for the GAA model with RMSE of 0.0791 outperformed MAE for the GAA model with RMSE of 0.0873.

The final step in modeling with the GP and GAA methods differs for each method. In selecting the best GP model, four basic arithmetic operators \((+,-,\times,\div)\) and the mathematical functions (sqrt, power, sin, cos, exp, abs) were employed. When input combination (2) and MAE as the fitness function were used, four different structures were chosen to identify the best GP model. The principal range of the investigated functions is:

\[
F1 = (+, -, \times, \div) \\
F2 = (+, -, \times, \div, \sin, \cos) \\
F3 = (+, -, \times, \div, \sin, \cos, \abs, \sqrt{x}, \power) \\
F4 = (+, -, \times, \div, \sin, \cos, \abs, \sqrt{x}, \power, \exp)
\]

The results signify that function \(F2\) with RMSE of 0.0515 performs superior to functions \(F1, F3, \) and \(F4\) with RMSE of 0.0521, 0.0629 and 0.0699, respectively.

The final step in selecting the best GAA model involved identifying the best transfer function. Therefore, several functions were chosen and their abilities investigated. Figure 2 represents the comparison between the four transfer functions. In this figure, the GAA model with the logarithmic transfer function in both hidden and output layers and with RMSE of 0.0672 performed better than the models with other transfer function combinations.

The two models introduced were compared in order to identify the optimal one. Figure 3 illustrates the \(\%SF_w\) plot of the two presented models for all data. As seen in Figure 3, the GP model made more appropriate predictions than the GAA model. According to the fit line equations of the scatter plots (assuming the equation is \(y = a_1x + a_2\)) the \(a_1\) coefficient of the GP model is closer to 1 and the \(a_2\) coefficient is closer to 0 compared to the GAA model. It is obvious from these scatter plots that the GP model estimates are less scattered and closer to the exact line than those of the GAA model. Evidently, the GAA model estimates \(\%SF_w\) with lower accuracy than the GP model, but the adjusted \(R^2\) value of the GAA model is higher than the GP model. This is acceptable because the output data of the GAA model is nearer to the mean of the experimental outputs.

In this study, the aim of GP modeling was to identify the most appropriate program for modeling the prediction of \(\%SF_w\) in rectangular channels. The best model program (Figure 4) in the form of a Matlab code indicates that by selecting input combination (2), MAE as a fitness function and mathematical function \(F2\) were obtained.

The output program of the GP model as the best model and equations proposed by researchers reported in literature were compared. Figure 5 illustrates the scatter plots of predicted and observed \(\%SF_w\) for all data using the GP model and the five above-mentioned equations. According to Figure 5, the GP model program with adjusted \(R^2\) of 0.9652 has the best function compared with the other equations. The equation obtained by Knight (1984) indicated the most suitable performance (with adjusted \(R^2\) of 0.9396) compared to other equations. The equations expressed by Knight et al. (1984, 1994) and Seckin
et al. (2006) for smooth channels and Rhodes & Knight (1994) for smooth ducts predicted %SFw well. It is worth noting that these equations predicted some points with nearly perfect adjustment for high aspect ratio values (B/h > 5). It can be deduced that by increasing the aspect ratio (B/h), roughness is less effective on the %SFw. In the GP program, three important parameters were considered. However, with input combination (2) roughness was not perceived, although the model made predictions very close to experimental data. In Knight’s equation only two parameters were involved although roughness was considered, and the results obtained with this relation were superior to the other mentioned equations. Yet once again, the GP model produced better results than the other equations.

CONCLUSION

The efficiency of soft computing methods in predicting the %SFw in rectangular channels with rough boundaries was investigated in this study. GP and GAA models were developed in three steps. First, the effective parameters were selected and different input combinations were tested to select the optimum combinations. Then the best fitness and transfer functions were studied and the superior one was accepted for both models. Finally, after extending the two models, their ability to predict the percentage of shear force was examined. According to performance results, the GP model predicted %SFw more accurately than the GAA model. The results of the proposed model were also compared with five equations presented by other researchers. The GP model predicted %SFw with lower error (RMSE = 0.0515) and higher accuracy than the equations by Knight (1981), Knight et al. (1984, 1994), Seckin et al. (2006) and Rhodes & Knight (1994) with RMSE of 0.0642, 0.2413, 0.2327, 0.2424 and 0.2063 respectively. The equations
proposed for smooth rectangular channels and ducts predicted overestimated \(\%SF_{w}\) values. It can be deduced that the relations for smooth boundaries do not yield more accurate results for estimating \(\%SF_{w}\) in channels with rough boundaries except for high aspect ratio (when \(B/h > 5\)). The equation obtained by Knight (1981) produced the most appropriate results after the best GP model, as the equation was also presented for rectangular channels with rough boundaries.

**REFERENCES**


First received 15 January 2015; accepted in revised form 1 September 2015. Available online 15 September 2015.