Turkish J. Eng. Env. Sci.
36 (2012) , 121 – 138.

doi:10.3906/muh-1008-30

A study of friction factor formulation in pipes using artificial intelligence techniques and explicit equations

Farzin SALMASI[1,*], **Rahman KHATIBI**[2], **Mohammad Ali GHORBANI**[1]
[1] *Department of Water Engineering, Faculty of Agriculture, Tabriz University, Tabriz-IRAN*
e-mail: Salmasi@Tabrizu.ac.ir, Ghorbani@Tabrizu.ac.ir
[2] *Consultant Mathematical Modeler, Swindon-UNITED KINGDOM*
e-mail: rahman_khatibi@yahoo.co.uk

Received: 17.08.2010

Abstract

The hydraulic design and analysis of flow conditions in pipe networks are dependent upon estimating the friction factor, f. The performance of its explicit formulations and those of artificial intelligence (AI) techniques are studied in this paper. The AI techniques used here include artificial neural networks (ANNs) and genetic programming (GP); both use the same data generated numerically by systematically changing the values of Reynolds numbers, R_e, and relative roughness, $\varepsilon/_D$, and solving the Colebrook-White equation for the value of f by using the successive substitution method. The tests included the transformation of R_e and $\varepsilon/_D$ using a logarithmic scale. This study shows that some of the explicit formulations for friction factor induce undue errors, but a number of them have good accuracy. The ANN formulation for the solving of the friction factor in the Colebrook-White equation is less successful than that by GP. The implementation of GP offers another explicit formulation for the friction factor; the performance of GP in terms of R^2 (0.997) and the root-mean-square error (0.013) is good, but its numerically obtained values are slightly perturbed.

Key Words: Pipe friction factor, Darcy-Weisbach equation, implicit/explicit equations, artificial neural network, genetic programming

1. Introduction

The understanding of flow equations in closed conduits reached its maturity in the early 20th century, whereby flows in such systems are driven by pressure differences between 2 different locations and the equation is referred to as the Darcy-Weisbach equation. This hydraulic equation serves as the basis for hydraulic design and analysis of water distribution systems, and it is expressed in terms of pressure drop, which is a directly measurable quantity of friction. However, the mathematical formulation of the problem includes an empirical friction parameter, f, for which the Colebrook-White equation is one well-known implicit formulation, such that the factor appears on both sides of the equation. This implicit problem is not intractable, as it can be

*Corresponding author

treated by iterative techniques, although it is cumbersome. Until the wide application of artificial intelligence (AI) in the 1990s, the challenge was to develop its explicit formulations, but, since then, the application of AI techniques is also a focus of research.

The Colebrook-White equation integrates important theoretical work by von Karman and Prandtl by accounting for both smooth and turbulent flow regimes in terms of 2 parameters, the Reynolds number, R_e, and the relative roughness as a measure of friction, ε/D. Alternative methods of solving the Colebrook-White equation include iterative methods, analytical solutions using the Lambert W function, use of an explicit equation, soft computing techniques that recognize that f-values are not precise, and a host of AI techniques used in recent years, including the artificial neural network (ANN) technique and genetic programming (GP). However, ANNs and GP assume that f-values are precise.

ANNs are parallel information processing systems that emulate the working processes in the brain. A neural network consists of a set of neurons or nodes arranged in layers; in the case that weighted inputs are used, these nodes provide suitable inputs by conversion functions (Kişi, 2005). Each neuron in a layer is connected to all of the neurons of the next layer, but without any interconnection among neurons in the same layer. Applications of ANNs to hydraulics go back to the 1990s and remain in active use today.

The GP methods, first proposed by Koza (1992), are wide-ranging and similar to genetic algorithms (Goldberg, 1989). GP techniques are robust applications of optimization algorithms and represent one way of mimicking natural selection. These techniques derive a set of mathematical expressions to describe the relationship between the independent and dependent variables using such operators as mutation, recombination (or crossover), and evolution. These are operated in a population evolving over generations through a definition of fitness and selection criteria. Applications of GP suit a wide range of problems and are particularly applicable to cases in which the interrelationships among the relevant variables are poorly understood or suspected to be wrong, or conventional mathematical analyses are constrained by restrictive assumptions but approximate solutions are acceptable (Banzhaf et al., 1998).

In smooth pipes, friction factor f depends only on R_e, and Gulyani (1999) provided a revision and discussion of the correlations more commonly used to estimate its value. However, the focus of recent research is largely on the full Colebrook-White equation.

More (2006) applied an analytical solution for the Colebrook-White equation for the friction factor using the Lambert W function. A close match was then observed by comparing the friction factor obtained from the Colebrook-White equation (used iteratively) and that obtained from the Lambert W function. Fadare and Ofidhe (2009) studied the ANN technique for the estimation of the friction factor in pipe flows and reported a high correlation factor of 0.999.

Yang et al. (2003) used ANNs to predict phase transport characteristics in high-pressure, 2-phase turbulent bubbly flows. Their investigation aimed to demonstrate the successful use of neural networks in the real-time determination of 2-phase flow properties at elevated pressures. They established 3 back-propagation neural networks, trained with the simulation results of a comprehensive theoretical model, to predict the transport characteristics (specifically the distributions of void-fraction and axial liquid-gas velocities) of upward turbulent bubbly pipe flows at pressures in the range of 3.5-7.0 MPa. Comparisons of the predictions with the test target vectors indicated that the root-mean-square error (RMSE) for each of the 3 back-propagation neural networks was within 5% to 6%.

To date, the application of GP in hydraulic engineering has been limited. Davidson et al. (1999)

determined empirical relationships for the friction in turbulent pipe flows and the additional resistance to flow induced by flexible vegetation, respectively. Giustolisi (2004) determined the Chezy resistance coefficient in corrugated channels. The authors are not aware of the application of GP to the Colebrook-White equation.

This paper is focused on treating the friction factor as having a precise value, but, in reality, this parameter is variable over time and data are often insufficient, ambiguous, and/or uncertain for precise treatments. Therefore, Yıldırım and Özger (2009), Yıldırım (2009), and Özger and Yıldırım (2009) investigated this parameter with soft computing techniques using various formulations of the friction factor, allowing them to identify precise values of friction values using neuro-fuzzy techniques.

The overall objective of the present study was to evaluate the performances of explicit formulations for estimating the friction factor, f, in the Darcy-Weisbach equation, while using ANNs and GP to avoid the need for a time-consuming and iterative solution of the Colebrook-White equation. The study involves the generation of data and comparisons between the various techniques with the numerical solutions of the Colebrook-White equation.

2. Models and methodology

2.1. Flow equation

The energy loss due to friction in Newtonian liquids flowing in a pipe is usually calculated with the Darcy-Weisbach equation, as follows.

$$h_f = f \frac{L}{D} \frac{V^2}{2g} \tag{1}$$

In this equation, f is referred to as the Moody or Darcy friction factor. This may be reformulated as follows.

$$f = \frac{D}{L} \frac{g\,h_f}{1/2\,V^2} = \frac{D}{L} \frac{\Delta P}{1/2\,\rho V^2} \tag{2}$$

The friction factor depends on the Reynolds number, R_e, and on the relative roughness of the pipe, ε/D.

For both smooth and turbulent flows, the friction factor is estimated with the following equation, developed by Colebrook and White (1937).

$$\frac{1}{\sqrt{f}} = -2\,\log\left(\frac{\varepsilon}{3.7D} + \frac{2.523}{R_e\sqrt{f}}\right) \tag{3}$$

2.2. Explicit formulations

The Colebrook-White equation is valid for R_e values ranging from 2000 to 10^8, and for values of relative roughness ranging from 0.0 to 0.05. The formula is often used in pipe network simulations. Its form is notably implicit, as the value of f appears on both sides of the equation, and its accurate solution is often very time-consuming, requiring many iterations. An approximate equation for f that does not require iteration can be used to improve the speed of simulation software. This was a subject of active research in the past, leading to a range of explicit formulations that are summarized in Appendix I. A study of the performance of these explicit equations was one of the aims of this paper.

2.3. Data specification and implementations of the AI models

The data in this modeling study were generated using a numerical procedure based on Eq. (3). The data generation included a systematic variation of R_e ranging from 2000 to 10^8 (using 74 values of R_e) and the varying of $^{\varepsilon}/_{D}$, ranging from 10^{-6} to 0.05 (28 values). Different combinations of R_e and $^{\varepsilon}/_{D}$ serve for the generation of data points in terms of f, where the f-value for each set of data is calculated by the numerical solution of the Colebrook-White equation using the successive substitution method. The dataset consisted of a total of 2072 points. Input variables were $^{\varepsilon}/_{D}$ and R_e, and the output was f. A selection of the generated data is shown in Appendix II and Table II.1.

2.4. Artificial neural networks

Any layer consists of predesignated neurons, and each neural network includes one or more of these interconnected layers. Figure 1 shows a 3-layered structure that consists of 1 input layer, I; 1 hidden layer, H; and 1 output layer, O. All of the neurons within a layer act synchronously. The operation process of these networks is such that the input layer accepts the data and the intermediate layers processes them, and, finally, the output layer displays the resulting outputs of the model application. During the modeling stage, the coefficients related to the present errors in the nodes are corrected by comparing the model outputs with the recorded input data (Rakhshandehroo et al., 2010).

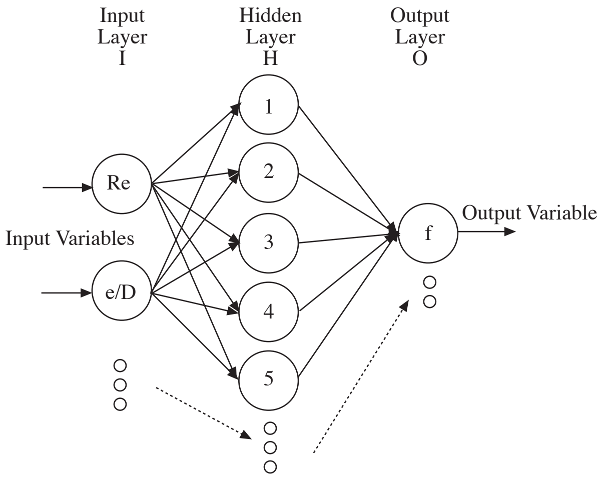


Figure 1. Neuron layout of artificial neural network (ANN).

The data for training the ANN model were generated using the numerical procedure described above. The dataset consisted of a total of 2072 points, of which 70% (1450 data points) were selected for the training process and 30% were selected as test data (622 data points). The optimal ANN configuration was selected from among various ANN configurations based on their predictive performances. The performance of the various ANN configurations was studied with 2 error measures: the determination coefficient, R^2, and RMSE.

2.5. Genetic programming

Implementation of GP models involves a number of preliminary decisions, including the selection of a set of basic operators such as $\left\{+, -, *, /, \wedge, \surd, \log, \exp, \sin, \arcsin, ...\right\}$ to construct a function, such as the reconstruction of an explicit equation for f, traditionally expressed by Eq. (3). The GP modeling programs provide operators like crossover and mutation to the winners, "children," or "offspring" to emulate natural selection, in which crossovers are responsible for maintaining identical features from one generation to another but mutations cause random changes. The evolution starts from an initially selected random population of models, where relationship f between the independent and dependent variables is often referred to as the "model," the "program," or the "solution." The population is allowed to evolve from one generation to another by virtue of a selected fitness criterion, and new models replace the old ones in this evolutionary process by having demonstrably better performance.

The study was carried out using GeneXpro software (Ferreira, 2001a, 2001b), which uses a gene expression method. Although this has differences from GP, both are inspired by natural selection in principle. For more detail on the implementation of GP models, see Ghorbani et al. (2010).

In this study, 4 basic arithmetic operators (+, - , × , /) and some basic mathematical functions ($\sqrt{}$, log, and e^x) were used. Like with ANNs, input variables were ε/D and R_e, and the output was f. A large number of generations (5000) were tested. The performance of GP was studied with 3 error measures: R^2, RMSE, and relative error (RE), as defined in Section 2.6.

2.6. Performance measures

The study involved comparisons, for which 3 performance measures were used to highlight different aspects of the problem. These measures were RMSE, R^2 and absolute RE which is defined as follows

$$RE = |f_{True} - f_{Estimated}| / f_{True}$$

Here, f_{True} is calculated from the Colebrook-White equation by the successive substitution method, and $f_{Estimated}$ is the output value from the explicit, GP, or ANN models.

3. Results

3.1. Performances of explicit equations

The performances of explicit equations for the Colebrook-White equations (presented in Appendix I) was investigated by comparing them against the numerical solution of the Colebrook-White equation for f-values using the data created in this study and presented in Appendix II. The results are presented in Figure 2, and their performances are summarized in Table 1 by categorizing them into 3 sets, those having inadequate, adequate, and good performances, with methods falling into the latter categories being very successful.

3.2. Implementation of ANN

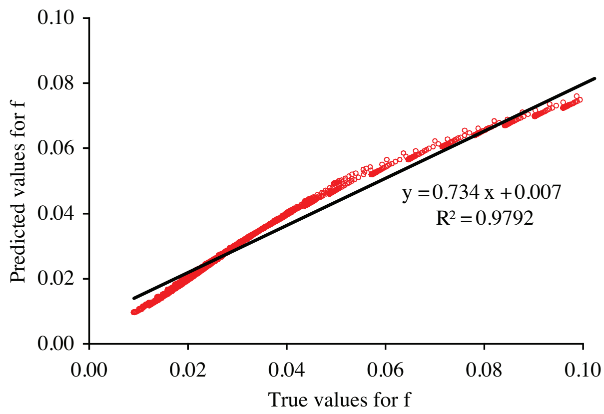
The initial identification of the model configurations did not employ any transformation of either data input or output. Its optimum configuration was selected by trial and error, by testing the set shown in Table 2. The identified architecture was 2-5-1 (input layer, 2 neurons; 1 hidden layer, 5 neurons; output layer, 1 neuron), for which the lowest RMSE was 0.0379 and the highest R^2 was 0.977. This led to inadequate predictions of f-values when the model was implemented in its prediction mode; in particular, unacceptably high errors were obtained for the predicted f-values corresponding to R_e values at the lower end of the chosen range. This indicated that the ANN model was unable to capture the initial curvature in the f-curves at the specified lower range of R_e.

Table 1. Error measurements in explicit equations, GP, and ANN with respect to numerical calculations.

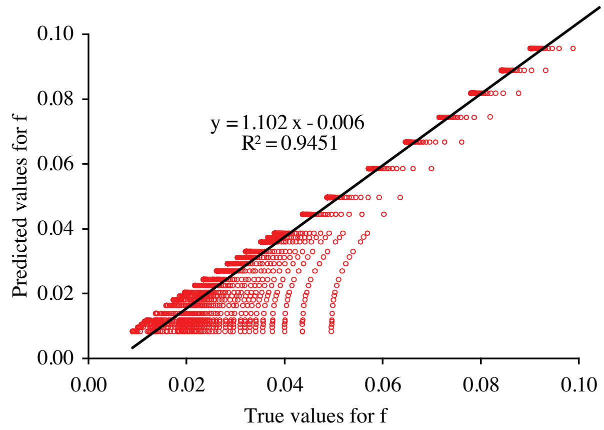
Study	$\hat{R}$	RMSE	Precision	Equation
Moody (1947)	0.9792	0.00787	Inadequate	I.4
Wood (1966)	0.9451	0.00735	Inadequate	I.5
Churchill (1977)	0.9669	0.00702	Inadequate	I.8
Churchill (1973)	0.9996	0.00062	Adequate	I.6
Swamee and Jain (1976)	0.9997	0.00055	Adequate	I.7
Barr (1981)	0.999991	0.0000792	Adequate	I.10
Chen (1979)	0.999996	0.0000617	Good performance	I.9
Zigrang and Sylvester (1982)	1.000000	0.0000213	Good performance	I.11
Manadilli (1997)	1.000000	0.0000102	Good performance	I.12
Romeo et al. (2002)	1.000000	0.0000092	Good performance	I.13
ANN	0.9951	0.0218	Locally inadequate	-
GP	0.9974	0.01324	Adequate	4
Color code for the performance of explicit equations				
Inadequate performance	Adequate performance	Good performance		

Table 2. Prediction errors for training and testing dataset of friction factor: different ANN configurations without transformations of the input parameters.

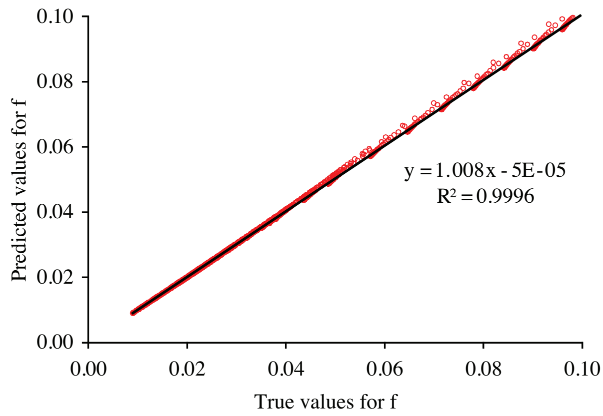
Transfer function	No. of hidden layers	No. of neurons/layer	Training RMSE	Training R^2	Test RMSE	Test R^2
Sigmoid	1	2	0.0384	0.978	0.0422	0.968
Sigmoid	1	3	0.0375	0.978	0.0406	0.974
Sigmoid	1	4	0.0383	0.977	0.0386	0.977
Sigmoid	**1**	**5**	**0.0379**	**0.977**	**0.0396**	**0.976**
Sigmoid	1	6	0.0388	0.977	0.0382	0.976
Sigmoid	1	8	0.0369	0.977	0.038	0.977
Sigmoid	1	10	0.0399	0.976	0.038	0.974
Hyperbolic Tangent	1	5	0.0372	0.978	0.0388	0.976
Gaussian	1	5	0.0404	0.974	0.0374	0.979
Sigmoid	2	2, 2	0.0385	0.977	0.0411	0.973
Sigmoid	2	2, 3	0.0401	0.974	0.0374	0.979



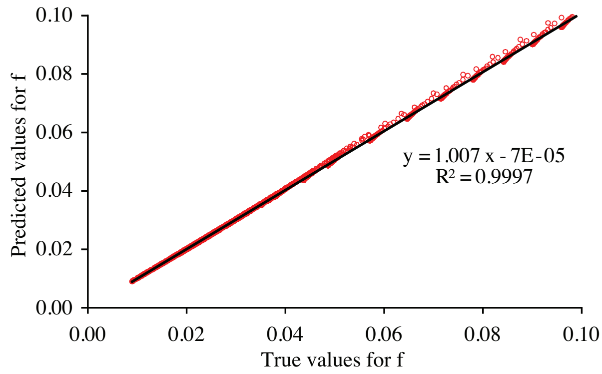
(a) Performance of explicit equation (Moody, 1947; Equation I.4).



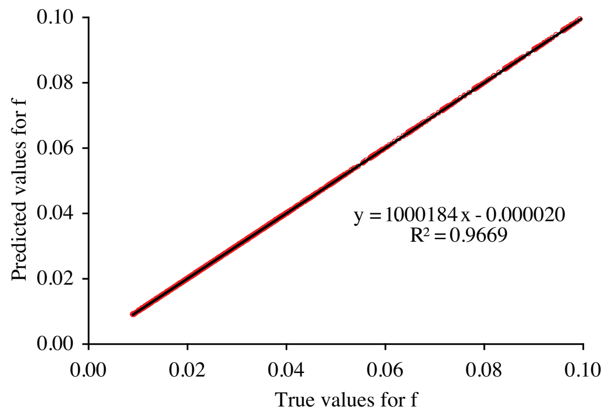
(b) Performance of explicit equation (Wood, 1966; Equation I.5).



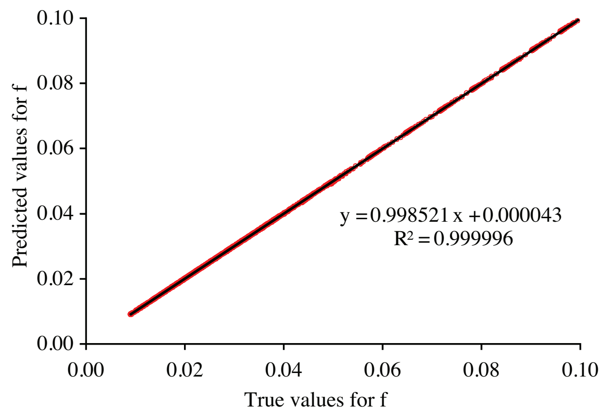
(c) Performance of explicit equation (Churchill, 1973; Equation I.6).



(d) Performance of explicit equation (Swamee and Jain, 1976; Equation I.7).

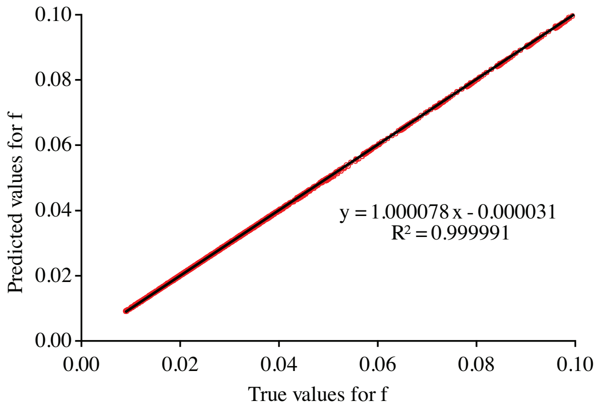


(e) Performance of explicit equation (Churchill, 1977; Equation I.8).

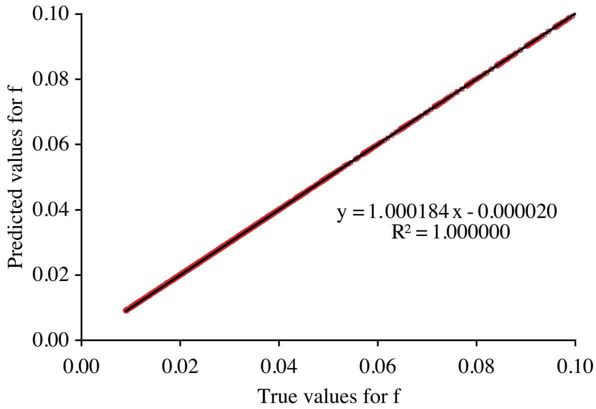


(f) Performance of explicit equation (Chen, 1979; Equation I.9).

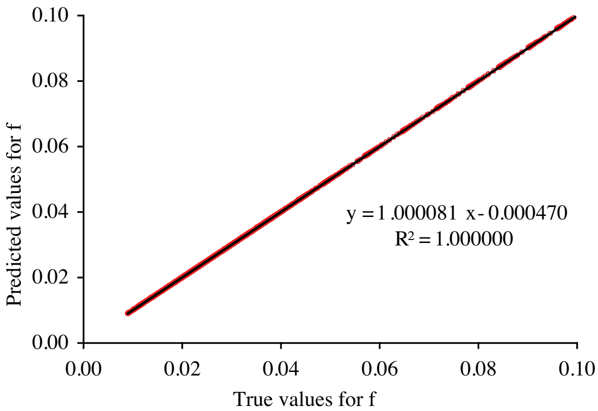
Figure 2. Performance of explicit equation against the numerical solution of the Colebrook-White equation in treating f-values.



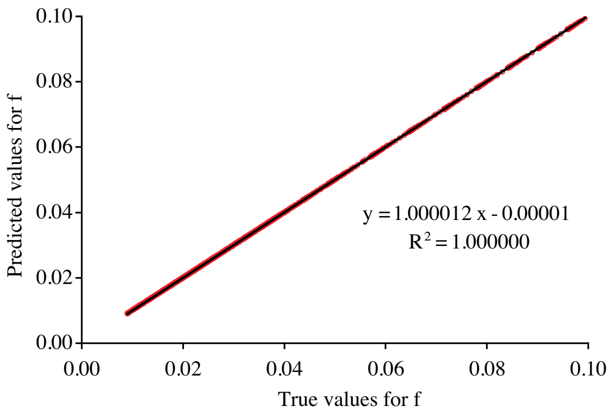
(g) Performance of explicit equation (Barr, 1981; Equation I.10).



(h) Performance of explicit equation (Zigrang and Sylvester, 1982; Equation I.11).



(i) Performance of explicit equation (Manadilli, 1997; Equation I.12).



(j) Performance of explicit equation (Romeo, et al. 2002; Equation I.13).

Figure 2. Continued.

Although ANN models do not require any prior knowledge of the relationships among inputs and outputs, a "warm start" is helpful to fine-tune the ANN model. For instance, it is clear from Eq. (3) that the parameter f is a logarithmic function of both input parameters, R_e and $\varepsilon/_D$. For this reason, another set of test runs were carried out to improve the performance of the ANN model by transforming both input data parameters. The R_e and $\varepsilon/_D$ parameters were transformed using a logarithmic function to the base of 10. The results, shown in Table 3, reveal that the optimum ANN configuration was improved markedly, as its RMSE was reduced to 0.0218 and its R^2 was increased to 0.995.

These results demonstrate the importance of choosing the right transformation of input data parameters and the significant impact that this may have on the overall performance of the ANN model.

3.3. Implementation of GP

The GP model was implemented by using the data in Appendix II. The functional setting and default parameters used in the GP modeling during this study are listed in Table 4. The GP model resulted in a highly nonlinear relationship with high accuracy and relatively low errors. The simplified analytic form of the proposed GP model may be expressed as follows.

Table 3. Predicted errors for training and testing dataset of friction factor; different ANN configurations with transformations of input parameters.

	No. of hidden	No. of	Training		Test	
Transfer function	layers	neurons/layer	RMSE	R^2	RMSE	R^2
Sigmoid	1	2	0.0325	0.981	0.0409	0.97
Sigmoid	1	3	0.0262	0.988	0.0266	0.987
Sigmoid	1	4	0.0353	0.989	0.0258	0.988
Sigmoid	**1**	**5**	**0.0218**	**0.995**	**0.0234**	**0.991**
Sigmoid	1	6	0.022	0.992	0.023	0.991

Table 4. Parameters of optimized GP model.

Parameter	Description of parameter	Setting of parameter
p1	Function set	+, −, ×, /, $\surd$, e^x, log
p2	Population size	250
p3	Mutation frequency (%)	96
p4	Crossover frequency (%)	50
p5	Number of replications	10
p6	Block mutation rate (%)	30
p7	Instruction mutation rate (%)	30
p8	Instruction data mutation rate (%)	40
p9	Homologous crossover (%)	95
p10	Program size	Initial 64, maximum 256

$$f = -0.0575 + \varepsilon/D + e^{-11.764(\varepsilon/D)-\log(2R_n)} + e^{-2.567+9.065/R_n-\varepsilon/D} \quad (4)$$

Figure 3 shows the RE in contour-line scheme by using the GP model from Eq. (4). The whole dataset (2072 points) has a mean RE of 2.52×10^{-5}, a maximum RE of 0.000117, and a minimum RE of 2.64×10^{-12}. The contour lines in Figure 3 show that the RE in the GP model is greater only in the upper right part of the graph. This area corresponds to $^{\varepsilon}/_{D} =$ 0.03, 0.02, and 0.015, and R_e values between 10^7 and 10^9. In other areas, the RE for the GP model is low and performs satisfactorily enough for the friction factor estimation. The error statistics of the GP model show that its RMSE and R^2 are 0.013 and 0.997, respectively, compared to the ANN quantitative performance values of RMSE = 0.022 and R^2 = 0.995. Therefore, the prediction accuracy of the GP model is generally better than that of the ANN model.

4. Discussion of the results

Engineering practices for pipe systems require the calculation of head losses and flows, and a common practice is to embed iterative methods for the calculation of f-values in the computer programs. However, this study shows that some of the explicit methods perform well and may replace the Colebrook-White equation, particularly in manual calculations, which can rapidly calculate f-values for given values of $^{\varepsilon}/_{D}$ and R_e. The investigations here show a sharp contrast in the performance of the explicit equations when compared with one another, but the accurate ones are attractive.

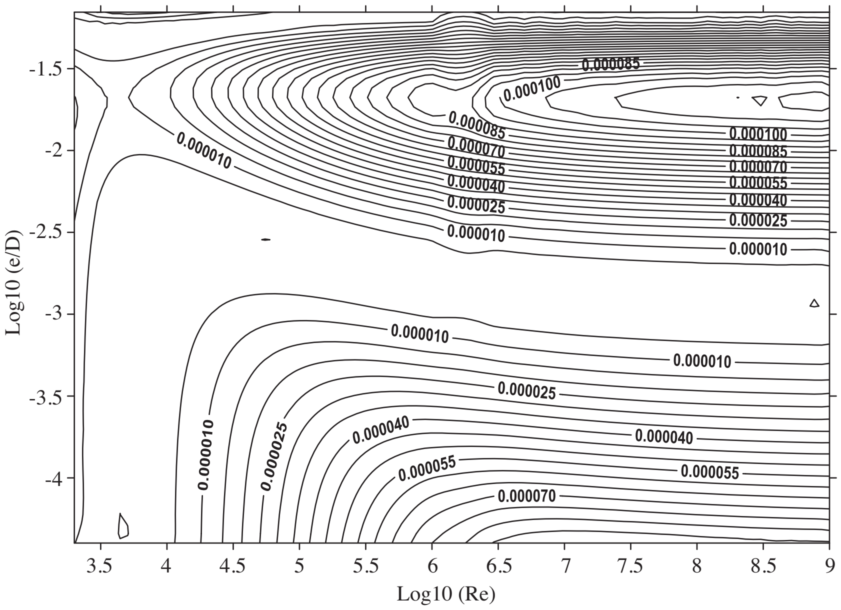


Figure 3. Contour of relative error for the GP model.

The performance of the ANN model in calculating the friction factor, f, was investigated by plotting a scatter diagram, as shown in Figure 4. Overall, the results were comparatively acceptable for calculating f, but the ANN model was less capable than some of the explicit equations, like those used by Chen (1979), Barr (1981), Zigrang and Sylvester (1982), and Romeo et al. (2002).

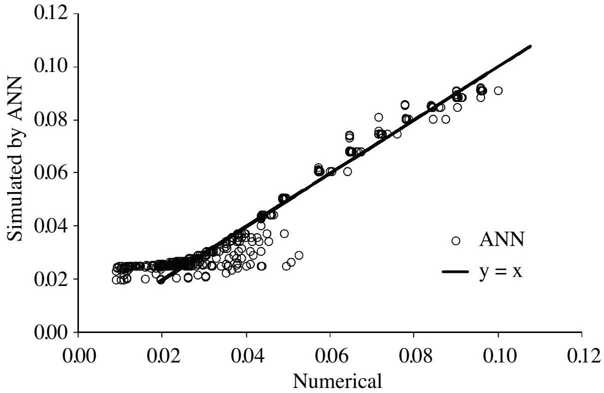


Figure 4. Scatter diagram for performance of ANN and numerical solution of the Colebrook-White equation.

The performance of the GP model in calculating friction factor f was investigated by plotting a scatter diagram, as shown in Figure 5. Overall, the GP model of the friction factor had some edge over the ANN model, both visually and quantitatively, but, at the same time, the GP model did not perform as well as some of the explicit formulations.

Future work will be directed toward improving the Colebrook-White equation for modern commercial pipes, like spiral and glass-reinforced plastic pipes.

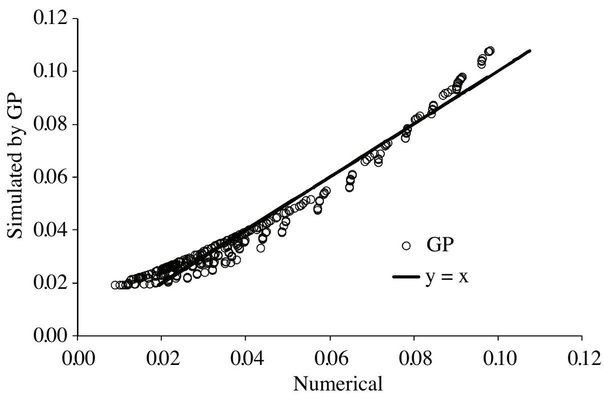


Figure 5. Scatter diagram for performance of GP and numerical solution of the Colebrook-White equation.

5. Conclusion

The paper focused on different methods used for predicting friction factor f in the Colebrook-White equation for calculating flows in pipes under pressure; the techniques selected were numerical solutions of the implicit Colebrook-White equations, various explicit forms of the Colebrook-White equation, and 2 applications of AI techniques, namely ANNs and GP techniques. The data were generated systematically for different values of the R_e and ε/D parameters using the Colebrook-White equation, and f-values were obtained using the successive substitution method for the equation's solution.

Preliminary test runs identified optimum ANN and GP models. The ANN model involved a neural network with 1 hidden layer and 5 neurons in that layer. Following the logarithmic transformations of the input data parameters, the trained network was able to perform better, with R^2 and RMSE values of 0.995 and 0.022, respectively (Table 3). The performance of the GP model using the testing data points showed a high generalization capacity, with $R^2 = 0.997$ and RMSE = 0.013. This model allows for an explicit solution of f without the need to employ a time-consuming iterative or trial-and-error solution scheme, an approach that is usually associated with the solution of the Colebrook equation in the turbulent flow regime of closed pipes.

Explicit equations remove the need for the iteration required for solving for the friction factor in the Colebrook-White equation, but this study shows that a number of them induce some undue errors. However, this study further identified some of the explicit formulations as accurate. The ANN formulation to solve for the friction factor in the Colebrook-White equation was less successful than the GP approach. Although the performance of GP in terms of R^2 and RMSE was good, its numerically obtained values were slightly perturbed, and the GP model did not perform as well as some of the explicit equations.

Appendix I

I.1. Explicit methods

The Colebrook-White equation is a formula often used in pipe network simulation software. Many explicit expressions have been developed to replace it, in which the value of f appears on both sides of the equation. These explicit formulations are approximations for f that do not require iteration, and they can hence be used to improve the speed of simulation software.

Table I.1. Error measurements in explicit equations, GP, and ANN with respect to numerical calculations.

No.	References	Mathematical expressions for the methods	Identifier	Applicability range
1	Prandtl and von Karman	$\frac{1}{\sqrt{f}} = -2\log\left(\frac{2.523}{R_e\sqrt{f}}\right)$	(I.1.a)	Smooth pipes: $\varepsilon = 0$
2		$\frac{1}{\sqrt{f}} = -2\log\left(\frac{\varepsilon / D}{3.71}\right)$	(I.1.b)	Fully developed turbulent flow
3	von Karman	$\frac{1}{\sqrt{f}} = -2\log(R_e\sqrt{f}) - 0.8 = 2\log\left(\frac{R_e\sqrt{f}}{2.523}\right)$	(I.2)	$R_e(\varepsilon / D)\sqrt{f} > 200$ f depends only on R_e
4	Method of successive substitution	$F_{n+1} = -2\log(\frac{\varepsilon}{3.7D} + \frac{2.523}{R_e}F_n)$; where F: $1/\sqrt{f}$	(I.3)	-
5	Moody (1947)	$f = 0.0055\left(1 + \left(2*10^4\frac{\varepsilon}{D} + \frac{10^6}{R_e}\right)^{\frac{1}{3}}\right)$	(I.4)	$4000 < R_e < 10^8$ $0.0 < \varepsilon/D < 0.01$
6	Wood (1966)	$f = a + b.R_e^{-c}$; where $a = 0.094\left(\frac{\varepsilon}{D}\right)^{0.225} + 0.53\left(\frac{\varepsilon}{D}\right)$ $b = 0.88\left(\frac{\varepsilon}{D}\right)$, $c = 1.62\left(\frac{\varepsilon}{D}\right)^{0.134}$	(I.5)	$4000 < R_e < 10^7$ $0.00001 < \varepsilon/D < 0.04$
7	Churchill (1973): using the transport model	$e^{-1/0.869\sqrt{f}} = \frac{\varepsilon / D}{3.7} + \left(\frac{7}{R_e}\right)^{0.9} \Leftrightarrow \frac{1}{\sqrt{f}} = -2\log\left(\frac{\varepsilon / D}{3.7} + \left(\frac{7}{R_e}\right)^{0.9}\right)$	(I.6)	-
8	Swamee and Jain (1976)	$f = \frac{1.325}{\left[Ln\left(\frac{\varepsilon}{3.7D} + \frac{5.74}{R_e^{0.9}}\right)\right]^2}$	(I.7)	$10^{-6} < \varepsilon/D < 0.05$ and $3 \times 10^3 < R_e < 10^8$

Table I.1. Continued.

No.	References	Mathematical expressions for the methods	Identifier	Applicability range
9	Churchill (1977)	$f=8\left(\left(\frac{8}{R_e}\right)^{12}+\left(A+B\right)^{-\frac{3}{2}}\right)^{\frac{1}{12}}$ $A=\left[-2\log\left(\left(\frac{\varepsilon/D}{3.7}\right)+\left(\frac{7}{R_e}\right)^{0.9}\right)\right]^{16}$ $B=\left(\frac{37530}{R_e}\right)^{16}$	(I.8)	Valid for the whole range of R_e (laminar, transition and turbulent)
10	Chen (1979)	$\frac{1}{\sqrt{f}}=-2\log\left(\frac{\varepsilon/D}{3.7065}-\frac{5.0452}{R_e}\log\left(\frac{(\varepsilon/D)^{1.1098}}{2.8257}+\frac{5.8506}{R_e^{0.8981}}\right)\right)$	(I.9)	Involves 2 iterations of Eq. (3) $4000<R_e<4\times10^8$ $0.0000005<\varepsilon/D<0.05$
11	Barr (1981): analogous to Chen (1979)	$\frac{1}{\sqrt{f}}=-2\log\left(\frac{\varepsilon/D}{3.7}-\frac{4.518\log\left(R_e/7\right)}{R_e\left(1+\left(R_e^{0.52}\left(\varepsilon/D\right)^{0.7}\right)\right)}\right)$	(I.10)	-
12	Zigrang and Sylvester (1982): similar to chen (1979), but with 3 iterations	$\frac{1}{\sqrt{f}}=-2\log\left(\frac{\varepsilon/D}{3.7}-\frac{5.02}{R_e}\log\left(\frac{\varepsilon/D}{3.7}-\frac{5.02}{R_e}\log\left(\frac{\varepsilon/D}{3.7}+\frac{13}{R_e}\right)\right)\right)$	(I.11)	-
13	Manadilli (1997): signomial-like equations	$\frac{1}{\sqrt{f}}=-2\log\left(\frac{\varepsilon/D}{3.70}-\frac{95}{R_e^{0.983}}-\frac{96.82}{R_e}\right)$	(I.12)	$5235<R_e<10^8$, any value of ε/D
14	Romeo et al. (2002)	$\frac{1}{\sqrt{f}}=-2\log\left(\frac{\varepsilon/D}{3.7065}-\frac{5.0272}{R_e}\log\left(\frac{\varepsilon/D}{3.827}-\frac{4.567}{R_e}\log\left(\left(\frac{\varepsilon/D}{7.7918}\right)^{0.9924}+\left(\frac{5.3326}{208.815+R_e}\right)^{0.9345}\right)\right)\right)$	(I.13)	$0<\varepsilon/D<0.05$ and $3\times10^3<R_e<1.5\times10^8$

Appendix II

Generation of data points for numerical study.

Table II.1. Sample of data with combinations of ε/D, R_e, and f.

Row	ε/D	R_e	f	Row	ε/D	R_e	f	Row	ε/D	R_e	f
1	0.00002	2000	0.04955	149	0.00006	2000	0.04958	1925	0.08	2000	0.09875
2	0.00002	3000	0.04361	150	0.00006	3000	0.04364	1926	0.08	3000	0.09600
3	0.00002	4000	0.03999	151	0.00006	4000	0.04003	1927	0.08	4000	0.09459
4	0.00002	5000	0.03747	152	0.00006	5000	0.03752	1928	0.08	5000	0.09373
5	0.00002	6000	0.03558	153	0.00006	6000	0.03563	1929	0.08	6000	0.09315
6	0.00002	7000	0.03408	154	0.00006	7000	0.03414	1930	0.08	7000	0.09273
7	0.00002	8000	0.03286	155	0.00006	8000	0.03292	1931	0.08	8000	0.09241
8	0.00002	9000	0.03184	156	0.00006	9000	0.03189	1932	0.08	9000	0.09217
9	0.00002	10,000	0.03096	157	0.00006	10,000	0.03102	1933	0.08	10,000	0.09197
10	0.00002	12,000	0.02952	158	0.00006	12,000	0.02958	1934	0.08	12,000	0.09167
11	0.00002	13,000	0.02891	159	0.00006	13,000	0.02898	1935	0.08	13,000	0.09156
12	0.00002	15,000	0.02788	160	0.00006	15,000	0.02796	1936	0.08	15,000	0.09138
13	0.00002	18,000	0.02664	161	0.00006	18,000	0.02672	1937	0.08	18,000	0.09118
14	0.00002	20,000	0.02596	162	0.00006	20,000	0.02605	1938	0.08	20,000	0.09108
15	0.00002	22,000	0.02537	163	0.00006	22,000	0.02546	1939	0.08	22,000	0.09099
16	0.00002	25,000	0.02460	164	0.00006	25,000	0.02470	1940	0.08	25,000	0.09090
17	0.00002	27,000	0.02416	165	0.00006	27,000	0.02426	1941	0.08	27,000	0.09084
18	0.00002	30,000	0.02357	166	0.00006	30,000	0.02367	1942	0.08	30,000	0.09077
19	0.00002	33,000	0.02305	167	0.00006	33,000	0.02316	1943	0.08	33,000	0.09072
20	0.00002	35,000	0.02274	168	0.00006	35,000	0.02286	1944	0.08	35,000	0.09069
21	0.00002	37,000	0.02245	169	0.00006	37,000	0.02257	1945	0.08	37,000	0.09066
22	0.00002	40,000	0.02206	170	0.00006	40,000	0.02219	1946	0.08	40,000	0.09062
23	0.00002	42,000	0.02182	171	0.00006	42,000	0.02195	1947	0.08	42,000	0.09060
24	0.00002	45,000	0.02148	172	0.00006	45,000	0.02162	1948	0.08	45,000	0.09057
25	0.00002	48,000	0.02118	173	0.00006	48,000	0.02132	1949	0.08	48,000	0.09055
26	0.00002	50,000	0.02099	174	0.00006	50,000	0.02113	1950	0.08	50,000	0.09053
27	0.00002	53,000	0.02072	175	0.00006	53,000	0.02087	1951	0.08	53,000	0.09051
28	0.00002	55,000	0.02055	176	0.00006	55,000	0.02070	1952	0.08	55,000	0.09050
29	0.00002	58,000	0.02032	177	0.00006	58,000	0.02047	1953	0.08	58,000	0.09048
30	0.00002	60,000	0.02017	178	0.00006	60,000	0.02033	1954	0.08	60,000	0.09047
31	0.00002	65,000	0.01982	179	0.00006	65,000	0.01999	1955	0.08	65,000	0.09045
32	0.00002	70,000	0.01951	180	0.00006	70,000	0.01969	1956	0.08	70,000	0.09043
33	0.00002	75,000	0.01923	181	0.00006	75,000	0.01941	1957	0.08	75,000	0.09041
34	0.00002	80,000	0.01897	182	0.00006	80,000	0.01916	1958	0.08	80,000	0.09040
35	0.00002	85,000	0.01873	183	0.00006	85,000	0.01893	1959	0.08	85,000	0.09038
36	0.00002	90,000	0.01851	184	0.00006	90,000	0.01871	1960	0.08	90,000	0.09037
37	0.00002	95,000	0.01831	185	0.00006	95,000	0.01851	1961	0.08	95,000	0.09036
38	0.00002	100,000	0.01812	186	0.00006	100,000	0.01833	1962	0.08	100,000	0.09035
39	0.00002	120,000	0.01746	187	0.00006	120,000	0.01769	1963	0.08	120,000	0.09032

Table II.1. Continued.

Row	ε/D	R_e	f	Row	ε/D	R_e	f	Row	ε/D	R_e	f
40	0.00002	150,000	0.01671	188	0.00006	150,000	0.01697	1964	0.08	150,000	0.09029
41	0.00002	180,000	0.01613	189	0.00006	180,000	0.01643	1965	0.08	180,000	0.09027
42	0.00002	200,000	0.01582	190	0.00006	200,000	0.01613	1966	0.08	200,000	0.09026
43	0.00002	250,000	0.01518	191	0.00006	250,000	0.01553	1967	0.08	250,000	0.09024
44	0.00002	300,000	0.01468	192	0.00006	300,000	0.01508	1968	0.08	300,000	0.09023
45	0.00002	350,000	0.01429	193	0.00006	350,000	0.01472	1969	0.08	350,000	0.09022
46	0.00002	400,000	0.01397	194	0.00006	400,000	0.01443	1970	0.08	400,000	0.09021
47	0.00002	450,000	0.01369	195	0.00006	450,000	0.01418	1971	0.08	450,000	0.09021
48	0.00002	500,000	0.01345	196	0.00006	500,000	0.01397	1972	0.08	500,000	0.09020
49	0.00002	600,000	0.01306	197	0.00006	600,000	0.01363	1973	0.08	600,000	0.09020
50	0.00002	700,000	0.01275	198	0.00006	700,000	0.01337	1974	0.08	700,000	0.09019
51	0.00002	800,000	0.01249	199	0.00006	800,000	0.01315	1975	0.08	800,000	0.09019
52	0.00002	900,000	0.01228	200	0.00006	900,000	0.01298	1976	0.08	900,000	0.09019
53	0.00002	1,000,000	0.01209	201	0.00006	1,000,000	0.01283	1977	0.08	1,000,000	0.09019
54	0.00002	3,000,000	0.01056	202	0.00006	3,000,000	0.01171	1978	0.08	3,000,000	0.09017
55	0.00002	5,000,000	0.01007	203	0.00006	5,000,000	0.01142	1979	0.08	5,000,000	0.09017
56	0.00002	8,000,000	0.00974	204	0.00006	8,000,000	0.01124	1980	0.08	8,000,000	0.09017
57	0.00002	10,000,000	0.00962	205	0.00006	10,000,000	0.01117	1981	0.08	10,000,000	0.09017
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
74	0.00002	1,000,000,000	0.00902	209	0.00006	1,000,000,000	0.01090	1998	0.08	1,000,000,000	0.09017
75	0.00004	2000	0.04956	223	0.00008	2000	0.04960	1999	0.09	2000	0.10416
76	0.00002	3000	0.04361	224	0.00008	3000	0.04366	2000	0.09	3000	0.10152
77	0.00004	4000	0.04001	225	0.00008	4000	0.04005	2001	0.09	4000	0.10017
78	0.00004	5000	0.03749	226	0.00008	5000	0.03754	2002	0.09	5000	0.09935
79	0.00004	6000	0.03560	227	0.00008	6000	0.03565	2003	0.09	6000	0.09880
80	0.00004	7000	0.03411	228	0.00008	7000	0.03416	2004	0.09	7000	0.09840
81	0.00004	8000	0.03289	229	0.00008	8000	0.03295	2005	0.09	8000	0.09810
82	0.00004	9000	0.03186	230	0.00008	9000	0.03192	2006	0.09	9000	0.09787
83	0.00004	10,000	0.03099	231	0.00008	10,000	0.03105	2007	0.09	10,000	0.09768
84	0.00004	12,000	0.02955	232	0.00008	12,000	0.02962	2008	0.09	12,000	0.09740
85	0.00004	13,000	0.02895	233	0.00008	13,000	0.02902	2009	0.09	13,000	0.09729
86	0.00004	15,000	0.02792	234	0.00008	15,000	0.02799	2010	0.09	15,000	0.09712
87	0.00004	18,000	0.02668	235	0.00008	18,000	0.02676	2011	0.09	18,000	0.09693
88	0.00004	20,000	0.02600	236	0.00008	20,000	0.02609	2012	0.09	20,000	0.09683
89	0.00004	22,000	0.02541	237	0.00008	22,000	0.02550	2013	0.09	22,000	0.09676
90	0.00004	25,000	0.02465	238	0.00008	25,000	0.02475	2014	0.09	25,000	0.09666
91	0.00004	27,000	0.02421	239	0.00008	27,000	0.02431	2015	0.09	27,000	0.09661
92	0.00004	30,000	0.02362	240	0.00008	30,000	0.02373	2016	0.09	30,000	0.09655
93	0.00004	33,000	0.02311	241	0.00008	33,000	0.02322	2017	0.09	33,000	0.09650
94	0.00004	35,000	0.02280	242	0.00008	35,000	0.02292	2018	0.09	35,000	0.09647
95	0.00004	37,000	0.02251	243	0.00008	37,000	0.02263	2019	0.09	37,000	0.09644
96	0.00004	40,000	0.02212	244	0.00008	40,000	0.02225	2020	0.09	40,000	0.09641
97	0.00004	42,000	0.02188	245	0.00008	42,000	0.02201	2021	0.09	42,000	0.09638

Table II.1. Continued.

Row	ε/D	R_e	f	Row	ε/D	R_e	f	Row	ε/D	R_e	f
98	0.00004	45,000	0.02155	246	0.00008	45,000	0.02169	2022	0.09	45,000	0.09636
99	0.00004	48,000	0.02125	247	0.00008	48,000	0.02139	2023	0.09	48,000	0.09633
100	0.00004	50,000	0.02106	248	0.00008	50,000	0.02120	2024	0.09	50,000	0.09632
101	0.00004	53,000	0.02079	249	0.00008	53,000	0.02094	2025	0.09	53,000	0.09630
102	0.00004	55,000	0.02063	250	0.00008	55,000	0.02078	2026	0.09	55,000	0.09629
103	0.00004	58,000	0.02040	251	0.00008	58,000	0.02055	2027	0.09	58,000	0.09627
104	0.00004	60,000	0.02025	252	0.00008	60,000	0.02040	2028	0.09	60,000	0.09626
105	0.00004	65,000	0.01991	253	0.00008	65,000	0.02007	2029	0.09	65,000	0.09624
106	0.00004	70,000	0.01960	254	0.00008	70,000	0.01977	2030	0.09	70,000	0.09622
107	0.00004	75,000	0.01932	255	0.00008	75,000	0.01950	2031	0.09	75,000	0.09620
108	0.00004	80,000	0.01906	256	0.00008	80,000	0.01925	2032	0.09	80,000	0.09619
109	0.00004	85,000	0.01883	257	0.00008	85,000	0.01902	2033	0.09	85,000	0.09618
110	0.00004	90,000	0.01861	258	0.00008	90,000	0.01881	2034	0.09	90,000	0.09617
111	0.00004	95,000	0.01841	259	0.00008	95,000	0.01861	2035	0.09	95,000	0.09616
112	0.00004	100,000	0.01822	260	0.00008	100,000	0.01843	2036	0.09	100,000	0.09615
113	0.00004	120,000	0.01758	261	0.00008	120,000	0.01781	2037	0.09	120,000	0.09612
114	0.00004	150,000	0.01684	262	0.00008	150,000	0.01710	2038	0.09	150,000	0.09609
115	0.00004	180,000	0.01628	263	0.00008	180,000	0.01657	2039	0.09	180,000	0.09607
116	0.00004	200,000	0.01597	264	0.00008	200,000	0.01628	2040	0.09	200,000	0.09606
117	0.00004	250,000	0.01535	265	0.00008	250,000	0.01570	2041	0.09	250,000	0.09604
118	0.00004	300,000	0.01488	266	0.00008	300,000	0.01526	2042	0.09	300,000	0.09603
119	0.00004	350,000	0.01451	267	0.00008	350,000	0.01492	2043	0.09	350,000	0.09602
120	0.00004	400,000	0.01420	268	0.00008	400,000	0.01464	2044	0.09	400,000	0.09602
121	0.00004	450,000	0.01394	269	0.00008	450,000	0.01441	2045	0.09	450,000	0.09601
122	0.00004	500,000	0.01372	270	0.00008	500,000	0.01421	2046	0.09	500,000	0.09601
123	0.00004	600,000	0.01336	271	0.00008	600,000	0.01389	2047	0.09	600,000	0.09600
124	0.00004	700,000	0.01307	272	0.00008	700,000	0.01365	2048	0.09	700,000	0.09600
125	0.00004	800,000	0.01284	273	0.00008	800,000	0.01345	2049	0.09	800,000	0.09600
126	0.00004	900,000	0.01264	274	0.00008	900,000	0.01329	2050	0.09	900,000	0.09599
127	0.00004	1,000,000	0.01248	275	0.00008	1,000,000	0.01315	2051	0.09	1,000,000	0.09599
128	0.00004	3,000,000	0.01119	276	0.00008	3,000,000	0.01216	2052	0.09	3,000,000	0.09598
129	0.00004	5,000,000	0.01083	277	0.00008	5,000,000	0.01191	2053	0.09	5,000,000	0.09598
130	0.00004	8,000,000	0.01059	278	0.00008	8,000,000	0.01176	2054	0.09	8,000,000	0.09598
131	0.00004	10,000,000	0.01051	279	0.00008	10,000,000	0.01171	2055	0.09	10,000,000	0.09598
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
148	0.00004	1,000,000,000	0.01014	283	0.00008	1,000,000,000	0.01149	2072	0.09	1,000,000,000	0.09597

References

Banzhaf, W., Nordin, P., Keller, P.E. and Francone, F.D., Genetic Programming, Morgan Kaufmann, San Francisco, CA, 1998.

Barr, D.I.H., "Solutions of the Colebrook-White Function for Resistance to Uniform Turbulent Flow", Proceedings of the Institution of Civil Engineers, 71, 529-536, 1981.

Chen, N.H., “An Explicit Equation for Friction Factor in Pipe”, Industrial and Engineering Chemistry Fundamentals, 18, 216-231, 1979.

Churchill, S.W., “Empirical Expressions for the Shear Stress in Turbulent Flow in Commercial Pipe”, AIChE Journal, 19, 375-376, 1973.

Churchill, S.W., “Friction Factor Equations Spans all Fluid Flow Regimes”, Chemical Engineering, 84, 91-102, 1977.

Colebrook, C.F. and White, C.M., “Experiments with Fluid Friction Roughened Pipes”, Proceedings of the Royal Society A, 161, 367-370, 1937.

Davidson, J.W., Savic, D.A. and Walters, G.A., “Method for Identification of Explicit Polynomial Formulae for the Friction in Turbulent Pipe Flow”, Journal of Hydroinformatics, 1, 115-126, 1999.

Fadare D.A. and Ofidhe, U.I., “Artificial Neural Network Model for Prediction of Friction Factor in Pipe Flow”, Journal of Applied Sciences Research, 5, 662-670, 2009.

Ferreira, C., “Gene Expression Programming in Problem Solving”, Invited Tutorial of the 6th Online World Conference on Soft Computing in Industrial Applications, 2001a.

Ferreira, C., “Gene Expression Programming: A New Adaptive Algorithm for Solving Problems”, Complex Systems, 13, 87-129, 2001b.

Ghorbani, M.A., Khatibi, R., Aytek, A. and Makarynskyy, O., “Sea Water Level Forecasting Using Genetic Programming and Comparing the Performance with Artificial Neural Networks”, Computers & Geosciences, 36, 620-627, 2010.

Giustolisi, O., “Using Genetic Programming to Determine Chezy Resistance Coefficient in Corrugated Channels”, Journal of Hydroinformatics, 6, 157-173, 2004.

Goldberg, D.E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989.

Gulyani, B.B., “Simple Equations for Pipe Flow Analysis”, Hydrocarbon Processing, 78, 67-78, 1999.

Kişi, Ö., “Daily River Flow Forecasting Using Artificial Neural Networks and Auto-Regressive Models”, Turkish Journal of Engineering and Environmental Sciences, 29, 9-20, 2005.

Koza, J.R., Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, MA, 1992.

Manadilli, G., “Replace Implicit Equations with Signomial Functions”, Chemical Engineering, 104, 129-132, 1997.

Moody, M.L., “An Approximate Formula for Pipe Friction Factors”, Transactions of the ASME, 69, 1005-1009, 1947.

More, A.A., “Analytical Solutions for the Colebrook and White Equation and for Pressure Drop in Ideal Gas Flow in Pipes”, Chemical Engineering Science, 61, 5515-5519, 2006.

Özger M. and Yıldırım G., “Determining Turbulent Flow Friction Coefficient using Adaptive Neuro-Fuzzy Computing Technique”, Advances in Engineering Software, 40, 281-287, 2009.

Rakhshandehroo, G.R., Vaghefi, M. and Shafiee, M.M., “Flood Forecasting in Similar Catchments Using Neural Networks”, Turkish Journal of Engineering and Environmental Sciences, 34, 57-66, 2010.

Romeo, E., Royo, C. and Monzon, A., “Improved Explicit Equations for Estimation of the Friction Factor in Rough and Smooth Pipes”, Chemical Engineering Journal, 86, 369-374, 2002.

Swamee, P.K. and Jain, A.K., “Explicit Equations for Pipe-Flow Problems”, Journal of the Hydraulics Division, 102, 657-664, 1976.

Wood, D.J., “An Explicit Friction Factor Relationship”, Civil Engineering, 36, 60-61, 1966.

Yang, A.S., Kuo, T.C. and Ling, P.H., “Application of Neural Networks to Prediction of Phase Transport Characteristics in High Pressure Two-Phase Turbulent Bubbly Flows”, Nuclear Engineering and Design, 223, 295-313, 2003.

Yıldırım G., “Computer-Based Analysis of Explicit Approximations to the Implicit Colebrook-White Equation in Turbulent Flow Friction Factor Calculation”, Advances in Engineering Software, 40, 1183-1190, 2009.

Yıldırım, G., and Özger, M., “Neuro-Fuzzy Approach in Estimating Hazen-Williams Friction Coefficient for Small-Diameter Polyethylene Pipes”, Advances in Engineering Software, 40, 593-599, 2009.

Zigrang, D.J. and Sylvester, N.D., “Explicit Approximations to the Solution of Colebrook’s Friction Factor Equation”, AIChE Journal, 28, 514-602, 1982.