Application of Artificial Intelligence to Obtain an Explicit Approximation of the Colebrook Equation

Fabrício M. Gomes¹,², Félix M. Pereira¹,², Ricardo B. Penteado¹,², Rubia F. T. Oliveira², Messias B. Silva¹,².
¹ Departamento de Engenharia Química – Escola de Engenharia de Lorena – Universidade de São Paulo (USP)
Lorena, SP, Brazil.
²Departamento de Produção – Faculdade de Engenharia de Guaratinguetá – Universidade Júlio Mesquita Filho (UNESP) – Guaratinguetá, SP, Brazil.
*fmgomes@usp.br

Abstract—This paper proposes a new explicit approximation of Colebrook equation using artificial intelligence by the Eureqa Analyzer software. The new approximation was compared with several models present in the literature. The results showed that, although the software did not obtain the best approximation compared to those presented in the literature, it was able to construct a model with a maximum relative error less than 1%. This method can be useful to build equations in other problems including implicit equations.

Keywords— Colebrook, Artificial Intelligence, Symbolic Regression.

I. INTRODUCTION

The determination of the friction factor in pipes for turbulent flow is essential not only to pressure drop calculations in pipelines and heat exchangers [1] but also is needed for calculating the Nusselt number in turbulent tube flow [2] and [3].

Colebrook [4] developed the following implicit equation that combines experimental results of turbulent flow in smooth and rough pipes, as shown in (1).

\[
\frac{1}{\sqrt{f}} = -2 \log \left( \frac{2.51}{Re \sqrt{f}} + \frac{\varepsilon}{3.71} \right)
\]  (1)

In (1), f is the friction factor, Re is the Reynolds number and \( \varepsilon \) is the relative pipe roughness.

Equation (1) was plotted in 1944 by Moody [5] into what is now called the Moody chart for pipe friction. This graph is certainly the most used tool to estimate the friction factor [6].

Since (1) is an implicit equation on f, iterative methods should be used to solve it, however, these methods commonly have as limitations the initial estimation dependence and the computational cost [7].

An alternative to the f estimation is the use of approximated explicit equations. These equations are commonly obtained by the regression between the data from the numerical solution of (1) and a pre-specified model to estimates the model parameters in order to obtain the f with a maxim precision. The success of this method dependent on the researcher expertise in to propose a precise equation.

Other alternative to obtain explicit equations to the friction factor estimation is the use of symbolic regression applying evolutionary algorithm. In this case, the equation is automatically obtained without imposing a pre-defined equation. The main objective of this paper is to demonstrate an approximation of the Colebrook equation using symbolic regression and compare its performance with several proposed approaches.

II. SYMBOLIC REGRESSION

Symbolic regression [8] is a method for searching the space of mathematical expressions, while minimizing various error metrics. Unlike traditional linear and nonlinear regression methods that fit parameters to an equation of a given form, symbolic regression searches both the parameters and the form of equations simultaneously [9].

To perform the symbolic regression this paper applied the Eureqa Formulize software [10]. The detailed description of the Eureqa software algorithm is presented on literature [10,11].

Recently, symbolic regression has been applied to several research areas, including chemistry [12], thermodynamics [13] and mechanical [14].
<table>
<thead>
<tr>
<th>Equation</th>
<th>Ref</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f = 0.0055 \left[ 1 + \left( \frac{20000 \cdot \varepsilon + 10^6}{Re} \right)^{1/3} \right] )</td>
<td>[2]</td>
<td>(2)</td>
</tr>
<tr>
<td>( f = 0.53 \cdot \varepsilon + 0.094 \cdot \varepsilon^{0.225} + 88 \cdot \varepsilon^{0.44} \cdot \text{Re}^{-1.62} \cdot \varepsilon^{1.14} )</td>
<td>[15]</td>
<td>(3)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{3.7 + 5.74}{\varepsilon} \right) )</td>
<td>[16]</td>
<td>(4)</td>
</tr>
<tr>
<td>( f = \left[ 1.14 - 2\log \left( \frac{21.25}{\varepsilon} \right) \right]^2 )</td>
<td>[17]</td>
<td>(5)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{3.7 + 5.74}{\varepsilon} \right) )</td>
<td>[18]</td>
<td>(6)</td>
</tr>
<tr>
<td>( f = \left[ -2\log \left( \frac{5.0452}{\varepsilon} \right) \right]^2 )</td>
<td>[19]</td>
<td>(7)</td>
</tr>
<tr>
<td>( f = -1.8\log \left( 0.135 \cdot \varepsilon + 6.5 \right) )</td>
<td>[20]</td>
<td>(8)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{4.518 \log(\text{Re})}{7} \right) ) \quad \begin{align} &amp;+ \left( \frac{\varepsilon}{3.7} \right) \end{align}</td>
<td>[21]</td>
<td>(9)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{5.02 \log(\text{Re})}{13} \right) ) \quad \begin{align} &amp;+ \left( \frac{\varepsilon}{3.7} \right) \end{align}</td>
<td>[22]</td>
<td>(10)</td>
</tr>
<tr>
<td>( f = -1.8\log \left( \frac{3.7 + 6.9}{\varepsilon} \right) )</td>
<td>[23]</td>
<td>(11)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{95 \cdot \varepsilon}{96.82 \cdot \text{Re}} \right) )</td>
<td>[24]</td>
<td>(12)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{5.0272 \log(\text{Re})}{3.827} \right) ) \quad \begin{align} &amp;- \frac{4.567}{\text{Re}} \end{align} \quad \begin{align} &amp;\log \left( \frac{\varepsilon}{7.79} \right) \quad \begin{align} &amp;+ \left( \frac{5.3326}{208.82 + \text{Re}} \right) \end{align} \quad \begin{align} &amp;0.9924 \end{align} \quad \begin{align} &amp;0.9345 \end{align}</td>
<td>[25]</td>
<td>(13)</td>
</tr>
<tr>
<td>( f = \left[ 0.8686 \log \left( \frac{0.4587 \cdot \text{Re}}{G^{0.50-10}} \right) \right]^2 )</td>
<td>[26]</td>
<td>(14)</td>
</tr>
<tr>
<td>where G is:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( G = 0.124 \cdot \text{Re} \cdot \varepsilon + \ln(0.4587 \cdot \text{Re}) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f = 1.613 \left[ \ln \left( \frac{0.234 \cdot \varepsilon^{1.1057} - 60.525}{\text{Re}^{1.1056} + 56.291} \right) \right]^2 )</td>
<td>[27]</td>
<td>(15)</td>
</tr>
<tr>
<td>( f = -2\log \left( \frac{2.18 \cdot \beta}{\text{Re}} + \frac{\varepsilon}{3.71} \right) )</td>
<td>[28]</td>
<td>(16)</td>
</tr>
<tr>
<td>where ( \beta ) is:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta = \ln \left( \frac{\text{Re}}{1.816 \cdot \ln \left( \frac{1.1 \cdot \text{Re}}{1 + 1.1 \cdot \text{Re}} \right)} \right) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE I. VARIOUS APPROXIMATIONS OF THE COLEBROOK’S EQUATION (CONTINUATION)

<table>
<thead>
<tr>
<th>Equation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f = \left( \frac{Re^2 - 0.6315093}{Re^{1/2} + Re^{-2\varepsilon}} \right) + 0.0275308 \left( \frac{6.929841 + \varepsilon}{Re} \right)^{1/6} )</td>
<td>[29]</td>
</tr>
<tr>
<td>( f = \left( \frac{2.51/Re + 1.1513 \cdot \delta}{\delta - (\varepsilon D/3.71 - 2.3026 \cdot \delta \cdot \log(\delta))} \right)^{1/2} )</td>
<td>[30]</td>
</tr>
</tbody>
</table>

where \( \delta = \frac{6.0173}{Re} \left( 0.07 \cdot (\varepsilon) + Re^{-0.109} \right) + \frac{\varepsilon}{3.71} \)

I. EXPLICIT EQUATIONS FOR CALCULATION OF THE FRICTION FACTOR

Table I shows the most widely used explicit approximations for the Colebrook’s equation postulated since 1947, in the order of publication year.

It is possible that there are other approaches present in the literature that are not present in this work. However, the greatest number of approximations were sought in order to ensure a meaningful comparison of the results.

II. METHOD

A. Eureqa Software configuration

Eureqa software has been programmed to work with 100% of the points for validation and training. The set points utilized in modeling is:

The total time for convergence was 4 hours

B. Models assessment

Several parameters can be used to the models assessment. In this study the following statistic criteria were used: mean relative error (MRE), maximal relative error (MXRE) and standard deviation of relative error (STRE). Especially, both the mean relative error and maximal relative error are very useful parameters for practically evaluating the most accurate model for friction factor estimation. The MRE calculation was performed by (19):

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f_{ij} - f_{pred,ij}}{f_{pred,ij}} \right| \times 100
\]  

The MXRE calculation was performed by (20):

\[
MXRE = \max \left( \frac{f_{ij} - f_{pred,ij}}{f_{ij}} \right) \times 100
\]  

The STRE calculation was performed by (21):

\[
STRE = \sqrt{\frac{\sum_{i=1}^{n} (f_{ij} - f_{pred,ij})^2}{n}}
\]  

In this paper, we will use the range of \( Re = 4000 - 10^8 \) a net will be formed using logarithm scale.

The way the points used to promote the accuracy of each approach were generated is described in Table II.

<table>
<thead>
<tr>
<th>Range</th>
<th>Nods</th>
<th>Linear Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>I ( Re = 4000 - 10^4 )</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>II ( Re = 10^4 - 10^5 )</td>
<td>90</td>
<td>1000</td>
</tr>
<tr>
<td>III ( Re = 10^5 - 10^6 )</td>
<td>90</td>
<td>10(^4)</td>
</tr>
<tr>
<td>IV ( Re = 10^6 - 10^7 )</td>
<td>90</td>
<td>10(^5)</td>
</tr>
<tr>
<td>V ( Re = 10^7 - 10^8 )</td>
<td>90</td>
<td>10(^6)</td>
</tr>
</tbody>
</table>

The relative pipe roughness (\( \varepsilon \)) values used were: 0, 1x10\(^{-6}\), 5x10\(^{-6}\), 1x10\(^{-5}\), 5x10\(^{-5}\), 1x10\(^{-4}\), 2x10\(^{-4}\), 4x10\(^{-4}\), 6x10\(^{-4}\), 8x10\(^{-4}\), 1x10\(^{-3}\), 2x10\(^{-3}\), 4x10\(^{-3}\), 6x10\(^{-3}\), 8x10\(^{-3}\), 1x10\(^{-2}\), 1.5x10\(^{-2}\), 2x10\(^{-2}\), 3x10\(^{-2}\), 4x10\(^{-2}\) and 5x10\(^{-2}\). These values were chosen because they are the same as those in the Moody chart [5].

The tests were then performed using 8,820 data points in total, and all statistic calculation were performed using Scilab v. 5.5.2.
The efficiency of the equation 

The use of artificial intelligence proved to be efficient in the search of an explicit equation for the friction factor estimation from data obtained by the Colebrook equation.

ACKNOWLEDGMENT

The authors would like to thank the Nutonian for giving academic license Eureqa Analyzer software.

REFERENCES


