Pet. Sci. (2015) 12:492–500
DOI 10.1007/s12182-015-0032-y

CrossMark

ORIGINAL PAPER

Model inference using the Akaike information criterion for turbulent flow of non-Newtonian crude oils in pipelines

Ahmed H. Kamel[1] · Ali S. Shaqlaih[2] · Essam A. Ibrahim[1]

Received: 8 December 2014 / Published online: 9 July 2015

Abstract The friction factor is a crucial parameter in calculating frictional pressure losses. However, it is a decisive challenge to estimate, especially for turbulent flow of non-Newtonian fluids in pipes. The objective of this paper is to examine the validity of friction factor correlations adopting a new informative-based approach, the Akaike information criterion (AIC) along with the coefficient of determination (R^2). Over a wide range of measured data, the results show that each model is accurate when it is examined against a specific dataset while the El-Emam et al. (Oil Gas J 101:74–83, 2003) model proves its superiority. In addition to its simple and explicit form, it covers a wide range of flow behavior indices and generalized Reynolds numbers. It is also shown that the traditional belief that a high R^2 means a better model may be misleading. AIC overcomes the shortcomings of R^2 as a trade between the complexity of the model and its accuracy not only to find a best approximating model but also to develop statistical inference based on the data. The authors present AIC to initiate an innovative strategy to help alleviate several challenges faced by the professionals in the oil and gas industry. Finally, a detailed discussion and models' ranking according to AIC and R^2 is presented showing the numerous advantages of AIC.

✉ Ahmed H. Kamel
kamel_a@utpb.edu

1 College of Business & Engineering, University of Texas of the Permian Basin, 4901 E University, Odessa, TX 79762-0001, USA

2 Department of Mathematics & Information Sciences, University of North Texas at Dallas, 7400 University Hills Blvd, Dallas, TX 75241, USA

Edited by Yan-Hua Sun

Keywords Friction factor · Pipeline · Information theory · Non-Newtonian · Turbulent

List of symbols

f	Fanning friction factor, dimensionless
i	Any models in the set
m	Number of models in the set
n	Flow behavior index, dimensionless
N_{Reg}	Generalized Reynolds number, dimensionless
R^2	Coefficient of determination
SSE	Summation of squared residuals
SSY	Summation of squared errors
ω	Akaike weight, dimensionless
Δ	Information lost compared with the best model
K	Number of the estimated parameters in the model
k	Consistency index of power law fluid, Pa s^n
v	Average fluid velocity, m/s
ρ	Fluid density, g/cc
K_p	Pipe consistency index, Pa s^n
E, φ, α	Parameters, in Eq. (7), function of flow behavior index
y_i	Any predicted data point
$\left(l\left(\hat{\theta}\vert y\right)\right)$	Nnumerical value of the likelihood at its maximum

1 Introduction

Throughout the world, large numbers of pipelines transport non-Newtonian pseudoplastic fluids including crude oils and petroleum products under turbulent flow conditions.

Pipeline design involves defining pipe size, thickness, and pumping head requirements where adequate provision for flow resistance is essential. The pumping head is required to overcome the inertia, gravity, and friction of the liquid inside the pipeline. The largest resistance force is the friction which is generally expressed in terms of pressure per unit length (dp/dl) required to overcome this resistance, psi/mile. The Darcy–Weisbach model is generally utilized to calculate the frictional pressure losses. Although it is a simple model, it involves a very essential factor; the friction factor. Numerous studies (Bogue 1962; Trinh 1969; Yoo 1974; Hanks and Ricks 1975; Govier 2008) have indicated that the friction factor is proportional to the kinetic energy of the fluid per unit volume and the area of the solid surface in contact with the fluid. This is the basis of the definition of the friction factor (Streeter and Wylie 1985). This friction factor or flow coefficient is not constant. Instead, it is defined in terms of the pipe specifications and fluid properties, but it is known to a high accuracy within certain flow regimes. For example, it was indicated that turbulent friction factors for non-Newtonian fluids can be obtained from the curves used for Newtonian fluids after the proper viscosity is inserted into the generalized Reynolds number (Govier 2008). However, there have been a number of advances in understanding the flow resistance of non-Newtonian fluids. As a result, many implicit and explicit equations; empirical, semi-empirical, and analytical have been proposed in the literature to accurately predict its value. Yet, they all seem to suffer from some drawbacks, either they are simple but not accurate or they are accurate but not simple.

So, the question is “which equation should be used?”. To answer this, a detailed comparative study among the published correlations is indispensable to select the best model while taking into consideration its simplicity. Data published by several authors including Dodge and Metzner (1959), Shaver and Merrill (1959), Yoo (1974), and Szilas et al. (1981) represent the basis of this comparison. The models involved in the comparison are selected based upon their accuracy, precision, simplicity, and range of applicability as indicated in the literature.

Previously, similar comparisons were based upon model selection methods; most commonly R^2. However, it is well documented that these methods still have some shortcomings (Anderson 2008; Shaqlaih 2010; Shaqlaih et al. 2013). For example, the coefficient of determination, R^2 is interpreted as an indication of the “goodness of fit” of the model. However, it may be misleading if data are associated with noise. Another interpretation of the R^2 coefficient is that the higher the coefficient of determination, the better the variance that the dependent variable is explained by the independent variable (Larson and Marx 2007). Yet, R^2 can be potentially increased by adding more independent variables to the model which makes it appear to be better while it is not. A third problem with this coefficient is that it does not give a clear indication on what value of R^2 should be used to categorize the good model versus the weaker model. There are many examples of models with relatively high R^2 but they do not represent good models (Burnham and Anderson 2002).

In this paper, a newly adopted technique in the oil and gas industry, based on the information theory approach (Akaike information criterion, AIC), is used. There are many reasons that make the AIC information theory a much better approach for model selection than many other well-known approaches. First, AIC is derived from the principles of information theory. Therefore, it models the information in the data rather than the data itself which are essential as data have noise (Claeskens and Hjort 2009; Shaqlaih et al. 2013). Second, AIC is theoretically sound as it is a mathematically derived formula not just a definition. The best model in this approach is the model that minimizes the information lost when the model is used to approximate the truth model (the perfect model to represent the data with the highest possible accuracy). Third, AIC penalizes the number of parameters in the model which means applying the parsimony principle in the model selection process and hence preventing over fitting (Burnham and Anderson 2002). Moreover, AIC gives a clear-cut way to distinguish between the poor models and the good models. In other words, AIC excludes the models that have a poor information-based representation of the truth model. Furthermore, it has been proven that the AIC approach is more stable in ranking the models than many other approaches (Shaqlaih 2010).

The analysis presented in this paper allows us to select the most precise model while not neglecting its simplicity. The authors believe that the application of the information theory approach and AIC will resolve various issues faced by oil and gas professionals related to model selection and will initiate an innovative strategy that has been demonstrated in other disciplines. Yet, it has not been used extensively in the oil and gas industry. The models, data, and analysis techniques are discussed thoroughly within the context of this paper.

2 Friction factor equations

The calculation of frictional pressure losses using the Darcy–Weisbach equation requires knowledge of the friction factor. It is worth recalling that the friction factor originally defined by Weisbach and Darcy friction factor is four times the Fanning friction factor (Moody 1944). It was shown, using dimensional analysis, that the friction factor and generalized Reynolds number are the two dimensionless groups obtainable from flow tests, and therefore, they are used to characterize the flow-resistance relationship. Many explicit and implicit equations have been proposed in the literature for the determination of the friction factor. The most common and accurate equations included in this study are explained in detail in the following paragraphs and then examined to check their validity and applicability. In this paper, the generalized Reynolds number can be given mathematically as:

$$N_{Reg} = \frac{d_i^n \rho v^{2-n}}{8^{n-1} K_p}. \tag{1}$$

2.1 Dodge and Metzner equation

Dodge and Metzner (1959) carried out a semi-theoretical analysis of the turbulent flow of time-independent fluids, pseudoplastic power law fluids, and applied techniques of dimensional analysis to derive their first semi-empirical formula for friction factors in circular pipes:

$$1\Big/\sqrt{f} = \frac{2}{n^{0.75}} \log\left(N_{Reg} f^{1-\frac{n}{2}}\right) - \frac{0.2}{n^{1.2}}. \tag{2}$$

They reported an excellent agreement between the calculated and the experimentally determined friction factors over a range of n from 0.36 to 1.0 and for generalized Reynolds numbers between 2900 and 100,000. The validity of this equation has been established for polymeric solutions, solid–liquid suspensions, power law, and non-power law fluids.

2.2 Shaver and Merrill equation

Shaver and Merrill (1959) developed a friction factor equation based on flow of aqueous plastic dispersions in smooth pipes:

$$f = \frac{0.79}{n^5 N_{Reg}^{\frac{2.62}{10.5n}}}. \tag{3}$$

It was reported that this empirical equation succeeded in correlating all the experimental data for n between 0.4 and 1.0. In fact, this correlation should not be used with n values lower than 0.4 due to its minimal accuracy (Shaver and Merrill 1959).

2.3 Tomita equation

Tomita (1959) developed his friction factor formula applying the Prandtle mixing length concept. The approximate validity of this equation was confirmed by 40 data points taken with starch pastes and lime slurries for flow behavior indices between 0.178 and 0.952 with generalized Reynolds numbers between 2000 and 100,000, as given below:

$$1\Big/\sqrt{f} = \sqrt{2}\log\left(N_{\mathrm{Reg}}\sqrt{\frac{f}{4}}\right) - \sqrt{0.2}. \tag{4}$$

2.4 Thomas equation

Thomas (1960) modified the Dodge and Metzner relationship to be given as:

$$1\Big/\sqrt{f} = \frac{\sqrt{2}}{n}\log\left(N_{Reg}\left(\frac{f}{4}\right)^{1-\frac{n}{2}}\right) - \frac{\sqrt{0.2}}{n}. \tag{5}$$

2.5 Clapp equation

Clapp (1961) applied the Prandtle and Von-Karman approach to derive a universal velocity profile and friction factor correlation for turbulent flow of power law fluids in smooth pipes. This equation reduces to an equation similar to the Nikuradse equation for $n = 1.0$, as following:

$$1\Big/\sqrt{f} = \frac{1.16}{n} - 1.22 + \frac{1.51}{n}\log\left(N_{Reg}\left(\frac{f}{4}\right)^{1-\frac{n}{2}}\right) + \frac{0.58}{n}(5n-8). \tag{6}$$

This equation was validated employing experimentally gathered data with a maximum deviation of ±4 % for $0.698 < n < 0.813$ and $548 < N_{Reg} < 42{,}800$.

2.6 Kemblowski and Kolodziejski equation

An alternative expression has been developed for the friction factor of power law fluids in turbulent flow (Kemblowski and Kolodziejski 1973). It is an empirical equation based on experimental data of aqueous suspensions with flow behavior indices ranging from 0.14 to 0.83 and generalized Reynolds numbers from 2680 to 98,600. E, φ, and m are defined elsewhere (Kemblowski and Kolodziejski 1973):

$$f = E \times \frac{\varphi^{\frac{1}{N_{Reg}}}}{N_{Reg}^{\alpha}}, \tag{7}$$

with

$$E = 0.0089 e^{3.57n^2}, \tag{7a}$$

$$\phi = e^{\frac{0.572\left(1-n^{4.2}\right)}{n^{0.0435}}}, \tag{7b}$$

$$\alpha = 0.314 n^{2.3} - 0.064. \tag{7c}$$

2.7 Garica and Steffe equation

Another equation was developed for the determination of the friction factor for pseudoplastic power law fluids (Garica and Steffe 1986) as follows:

$$1\Big/\sqrt{f} = 1.318\ln\left(N_{Reg}\sqrt{f} - 0.398\right). \quad (8)$$

2.8 Szilas et al. equation

A friction factor equation was developed by Szilas et al. (1981) as the first analytical relationship for flow of Non-Newtonian power law crude oils:

$$1\Big/\sqrt{f} = \frac{\sqrt{2}}{n}\log\left(N_{Reg}f^{1-\frac{n}{2}}\right) + 1.23^{\frac{1}{n}}\left(\frac{0.707}{n} + 2.12\right) - \frac{2}{n} - 1.028. \quad (9)$$

This equation was experimentally verified using data from the Hungarian Algyo crude oil pipeline for generalized Reynolds numbers varying between 10,000 and 100,000. This equation proved its accuracy when compared with several other equations (Szilas et al. 1981).

2.9 Desouky and El-Emam equation

Desouky and El-Emam (1990) derived an equation for designing a pipeline handling any type of pseudoplastic fluids under turbulent flow conditions by integrating the velocity distribution over the cross sectional area of the pipeline:

$$f = 0.71n^{n}\left(0.0112 + N_{Reg}^{-0.3185}\right). \quad (10)$$

A comparison with experimental data for pseudoplastic fluids measured by Yoo (1974) showed an excellent agreement with an average error of 2.6 % for all values of n (0.241 to 0.893).

2.10 Hawase et al. equation

Hawase et al. (1994) proposed an explicit expression for friction factor for hydraulically smooth pipes:

$$1\Big/\sqrt{f} = 1.89\log\left(\frac{N_{Reg}^{n^{\frac{1}{0.615}}}}{6.5n^{\frac{1}{1+0.75n}}}\right). \quad (11)$$

The values of f were within an error bound of ±2.4 % for $0.3 < n < 1$ and $4000 < N_{Reg} < 100{,}0000$ when compared with the predictions from the implicit expression of Dodge and Metzner.

2.11 El-Emam et al. equation

El-Emam et al. (2003) employed the data measured by several authors (Dodge and Metzner 1959; Shaver and Merrill 1959; Yoo 1974; Szilas et al. 1981) and developed a new empirical equation to calculate the friction factor for turbulent flow of non-Newtonian fluids. Their equation was statistically examined versus several other equations and experimental data and proved its accuracy:

$$f = \frac{n}{3.072 - 0.143n}N_{Reg}^{\frac{n}{0.282-4.211n}} - 0.00065. \quad (12)$$

Furthermore, the El-Emam et al. equation, in addition to several other equations, were evaluated using field data from an Egyptian pipeline; the Melieha-Al-Hamrah pipeline (101 miles long and 16-in. in diameter) which confirmed their proposed equation as a more realistic and simple approach (El-Emam et al. 2003).

Other equations are available in the literature as well (Torrance 1963; Trinh 1969; Hanks and Dadia 1971: Hanks and Ricks 1975; Derby and Melson 1981; Shenoy and Saini 1982; Shenoy 1988; Irvine 1988; Tam and Tiu 1988; Hemeida 1993; Trinh 2005). However, they are not included in the analysis due to either complexity, for example, they incorporate other dimensionless numbers such as the Hedstrom number, Deborah number, etc., or their limited validity when evaluated statistically or experimentally (Bogue 1962; Garica and Steffe 1986; Hartnet and Kostic 1990; Khaled 1994; El-Emam et al. 2003; Gao and Zhang 2007).

Table 1 lists the equations used in the present study along with their application ranges for flow behavior indices and generalized Reynolds numbers.

3 Measured data

The measured friction factors at different values of flow behavior indices and generalized Reynolds numbers incorporated in this analysis were gathered and published by several authors. Dodge and Metzner (1959) published friction factor values at flow behavior indices of 0.617, 0.726, and 1.0, while Shaver and Merrill (1959) published friction factor values at flow behavior indices of 0.6, 0.7, and 0.9. Other sets of data are published by Yoo (1974) at different values of flow behavior indices covering a wide range from 0.241 to 0.893 as well as Szilas et al. for $n = 0.5287$, 0.6991, 0.7169, 0.8311, and 0.948 (1981). The four sets of data are included in the analysis individually and collectively to cover a wide range of both generalized Reynolds number and flow behavior indices.

4 Model selection methods

Model selection methods refer to the criteria or strategy by which one can identify the most accurate model among a set of candidate models. However, there are many different strategies to select the best model from a set of candidate models. In this study, the widely used statistical procedure (R^2) method and the information theory approach, Akaike

Table 1 Fanning friction factor equations and application ranges

Model	Formula	Notes	Year
D & M	$1/\sqrt{f} = \frac{2}{n^{0.75}}\log\left(N_{Reg}f^{1-\frac{n}{2}}\right) - \frac{0.2}{n^{1.2}}$	$0.36 < n < 1.0$ $2900 < N_{Reg} < 100{,}000$	1959
S & M	$f = \frac{0.79}{n^5 N_{Reg}^{\frac{2.62}{10.5n}}}$	$0.4 < n < 1.0$ $4000 < N_{Reg} < 100{,}000$	1959
Tomita	$1/\sqrt{f} = \sqrt{2}\,\log\left(N_{\mathrm{Reg}}\sqrt{\frac{f}{4}}\right) - \sqrt{0.2}$	$0.178 < n < 0.952$ $2000 < N_{Reg} < 100{,}000$	1959
Thomas	$1/\sqrt{f} = \frac{\sqrt{2}}{n}\log\left(N_{Reg}\left(\frac{f}{4}\right)^{1-\frac{n}{2}}\right) - \frac{\sqrt{0.2}}{n}$	$0.36 < n < 1.0$ $2900 < N_{Reg} < 100{,}000$	1960
Clapp	$1/\sqrt{f} = \frac{1.16}{n} - 1.22 + \frac{1.51}{n}\log\left(N_{Reg}\left(\frac{f}{4}\right)^{1-\frac{n}{2}}\right) + \frac{0.58}{n}(5n-8)$	$0.698 < n < 0.813$ $548 < N_{Reg} < 42{,}800$	1961
K & K	$f = E \times \frac{\phi^{\frac{1}{N_{Reg}}}}{N_{Reg}^{\alpha}}$ $E = 0.0089e^{3.57n^2}$ $\phi = e^{\frac{0.572\left(1-n^{4.2}\right)}{n^{0.0435}}}$ $\alpha = 0.314n^{2.3} - 0.064$	$0.14 < n < 0.83$ $2{,}680 < N_{Reg} < 98{,}600$	1973
SBN	$1/\sqrt{f} = \frac{\sqrt{2}}{n}\log\left(N_{Reg}f^{1-\frac{n}{2}}\right) + 1.23^{\frac{1}{n}}\left(\frac{0.707}{n} + 2.12\right) - \frac{2}{n} - 1.028$	$0.24 < n < 1.0$ $10{,}000 < N_{Reg} < 100{,}000$	1981
G & S	$1/\sqrt{f} = 1.318\ln\left(N_{Reg}\sqrt{f} - 0.398\right)$	$0.4 < n < 0.82$ $3000 < N_{Reg} < 50{,}000$	1986
D & E	$f = 0.71n^n\left(0.0112 + N_{Reg}^{-0.3185}\right)$	$0.241 < n < 0.893$ $4000 < N_{Reg} < 100{,}000$	1990
HSW	$1/\sqrt{f} = 1.89\log\left(\frac{N_{Reg}^{\frac{1}{n^{0.615}}}}{6.5n^{\frac{1}{1+0.75n}}}\right)$	$0.3 < n < 1.0$ $4000 < N_{Reg} < 100{,}000$	1994
El-Emam et al.	$f = \frac{n}{3.072-0.143n}N_{Reg}^{\frac{n}{0.282-4.211n}} - 0.00065$	$0.178 < n < 1.0$ $4000 < N_{Reg} < 150{,}000$	2003

information criterion (AIC) are used. The coefficient of multiple determinations, R^2 for a model is defined as:

$$R^2 = 1 - \frac{SSE}{SSY} \tag{13}$$

where SSE and SSY are given as:

$$SSE = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{13a}$$

$$SSY = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \tag{13b}$$

where $\bar{y}$ is the average value of the observed values y_i, and $\hat{y}_i$ is predicted value of y_i under the model (Mendenhall and Sincich 2003). With the R^2 method, the larger the R^2, the more accurate the model (Mendenhall and Sincich 2003; Larson and Marx 2007). Even though R^2 is widely used as an indication of the goodness of fit of models, it should not be used with nonlinear models. However, it is used in this paper to prove that it is not a good measure for the models fit (Anderson 2008; Shaqlaih et al. 2013).

On the other hand, in the information theory approach, it is thought of the full reality as a model to be approximated (Burnham and Anderson 2002), and the objective is to find the model that best approximates the unknown truth model. Akaike (1973) showed that the model that best approximates the truth model is the one with smallest value of AIC:

$$AIC = -2\log\left(l\left(\hat{\theta}|y\right)\right) + 2K, \tag{14}$$

where K is the number of the estimated parameters in the model and $\left(l\left(\hat{\theta}|y\right)\right)$ is the numerical value of the likelihood at its maximum (Akaike 1973). The value of AIC gives the information lost if the chosen model is used to approximate the truth model. In other words in the information theory approach, the smaller the AIC, the more accurate the model. It is useful to define the AIC difference as: $\Delta_i = AIC_i - AIC_{\min}$, where $AIC_{\min}$ is the smallest value of the AIC values for all the set of candidate models. The best model has a Δ value of zero. A candidate model with Δ value higher than 10 should not be considered as a useful model (Anderson 2008; Shaqlaih et al. 2013). Another parameter is the Akaike's weight, ω_i:

$$\omega_i = \frac{\exp\left(-\frac{\Delta_i}{2}\right)}{\sum_{r=1}^{m} \exp\left(-\frac{\Delta_i}{2}\right)} \pi r^2, \tag{15}$$

where Δ_i is the AIC difference of the model i and m is the number of candidate models. ω_i gives the weight of evidence in favor of model i being the best model in the set of m models. One of the approaches to create a 95 % confidence set of models in the information theory approach is based on Akaike weights. In this approach, we sum the Akaike weights from largest to smallest until the sum is just $\geq$0.95. In the information theory approach, it is essential to find the Akaike weight for each model to be able to see the probability of the model being the best model. Akaike weight, AIC differences, and the confidence set of models are all essential tools in the model selection process in the information theory approach.

5 Results and discussion

In this study, both R^2 and the AIC are used to check for the best model among a set of the 11 candidate models discussed previously. For better understanding of the best model that approximates the friction factor, the published four sets of data are used individually. Later, these four sets are combined and used collectively as one set to examine the same models. The detailed results are discussed in the following paragraphs.

The first set of data was published by Dodge and Metzner (1959) for three different values of flow behavior index (0.617, 0.726, and 1.0) and a wide range of generalized Reynolds numbers. Table 2 shows the results for both R^2 and AIC methods. This table shows that according to R^2 values, Clapp (1961) and Desouky and El-Emam (1990) models are the best fit for the data with R^2 values of 0.92. Dodge and Metzner (1959) and El-Emam et al. (2003) models still have reasonable fit as their R^2 values are 0.88 and 0.85, respectively. Other models have reasonable R^2 values as well. Tomita (1959), Thomas (1960), and Garica and Steffe (1986) models have poor fits as each has R^2 value less than 0.50. As stated earlier, the R^2 method does not give a clear-cut evaluation of which models should be considered. For example, the Shaver and Merrill (1959) model has R^2 value of 0.66 which may be considered reasonably large. However, on the other hand, it is considerably less than 0.92, the largest R^2 value. The same conclusion applies to other models.

Regarding the AIC, as we can see in table, only two models can be used to accurately predict the friction factor; namely the Clapp (1961) and the Desouky and El-Emam models (1990). In fact the best model to use is the Clapp (1961) model with an Akaike weight, ω of 59.0 %. The Desouky and El-Emam (1990) model still has a recognizable Akaike weight of 41.0 %. We recall here that the Akaike weight provides evidence for which model is the best. The other models have no chance of being good models. Even though the Dodge and Metzner (1959) model was developed using this data, its performance is very unsatisfactory. The same results can be attained by looking at the Δ values. Indeed, we can see that the best model is the Clapp (1961) model with a Δ value of zero (or equivalently the smallest AIC value and hence the best model). The Desouky and El-Emam (1990) model is second best with a Δ-value of 0.75. We can clearly see that all other models should not be considered as their Δ-values are higher than 10 and hence their Akaike weights are negligible.

Table 2 Ranking of the correlations using Dodge and Metzner (1959) data

Model	R^2	R^2 ranking	Δ	ω	AIC ranking
D & M	0.88	2	14.3	0.00	Poor
S & M	0.66	6	51.8	0.00	Poor
Tomita	0.37	Poor	101.7	0.00	Poor
Thomas	0.17	Poor	83.5	0.00	Poor
Clapp	0.92	1	0.0	0.59	1
K & K	0.76	5	38.7	0.00	Poor
SBN	0.81	4	31.2	0.00	Poor
G & S	0.08	Poor	87.6	0.00	Poor
D & E	0.92	1	0.8	0.41	2
HSW	0.56	7	61.3	0.00	Poor
El-Emam et al.	0.85	3	21.6	0.00	Poor

However, AIC states that the Clapp model is better than Desouky and El-Emam (1990) model as the ratio between their Akaike weights is 0.59/0.41 = 1.4 which means that the Clapp model is 1.4 times better than the Desouky and El-Emam (1990) model, as inferred from the weight factor according to AIC definitions. The advantages of the AIC method over R^2 are clear as it gives the set of models that can be considered. Moreover, the AIC method not only ranks the models but also separates the models that should not be considered. Furthermore, the Akaike ratio clarifies how the selected models should be preferred (Burnham and Anderson 2002; Shaqlaih 2010).

Table 3 shows the results using Shaver and Merrill data for flow behavior indices of 0.6, 0.7, and 0.9. According to the R^2, Shaver and Merrill (1959) is the best model (highest R^2 value) which is logically true since this is the data used to develop the model. Also, same conclusion can be inferred for AIC.

Table 3 Ranking of the correlations using Shaver and Merrill (1959) data

Model	R^2	R^2 ranking	Δ	ω	AIC ranking
D & M	0.74	6	133.1	0.00	Poor
S & M	0.99	1	0.0	1.00	1
Tomita	0.12	11	183.7	0.00	Poor
Thomas	0.13	10	102.9	0.00	Poor
Clapp	0.81	4	119.7	0.00	Poor
K & K	0.44	9	164.6	0.00	Poor
SBN	0.89	2	94.9	0.00	Poor
G & S	0.71	8	174.5	0.00	Poor
D & E	0.76	5	129.9	0.00	Poor
HSW	0.74	7	133.2	0.00	Poor
El-Emam et al.	0.89	3	97.2	0.00	Poor

However, according to AIC all other models are weak (zero values for Akaike weights) and should not be considered, a result that could not be attained using R^2 values alone.

Similar conclusions can be generated using Yoo (1974) data and Szilas et al. (1981) data for other ranges of flow behavior indices and generalized Reynolds numbers.

Table 4 shows the results when using Yoo data while Table 5 shows similar results for Szilas et al. (1981) data. For the Yoo (1974) data in Table 4, both R^2 and AIC indicate that the Desouky and El-Emam (1990) model is the best in the set. Again, this is reasonably accepted since the Desouky and El-Emam (1990) showed that their model had an excellent agreement with an average error of 2.6 % for all the values of n when compared with the data measured by Yoo (1974). However, AIC analysis suggests that all other models should not be considered as they all have Akaike weights of 0.0.

Similarly, Table 5 shows that the Szilas et al. (1981) model is the best model when using Szilas data. Again, this conclusion is reasonably accepted. Even though El-Emam et al. (2003) and Thomas (1960) models have very high value of R^2 (0.99 and 0.97, respectively), they seem to be poor models according to AIC ranking. Recall that this is one of the disadvantages of the R^2 method (Burnham and Anderson 2002; Shaqlaih 2010).

In general, with each set of data, a specific model is believed to be the best either because it was developed using this set of data or because, when developed, it was compared and examined with this set of data to show its accuracy. Now, all four sets of data are combined and the same models are examined using R^2 and AIC. It is worth mentioning that combining all sets of data covers a very wide range of flow behavior indices n and generalized Reynolds numbers N_{Reg}. The analysis in this case is believed to be more realistic and the results should be statistically valid. The results are summarized in Table 6.

From Table 6, it can be seen that none of the previously selected models, for example, the Desouky and El-Emam (1990) model based on Yoo data and the Szilas et al. model based on Szilas data can predict accurate values of the friction factor for this wide range of n and N_{Reg} values. This may be due their application range. Most of these equations were empirically derived and experimentally verified with measured data covering a certain range of n and N_{Reg} values. Extending their application beyond this range is not normally possible and can lead to erroneous results. Using all data collectively showed that a different model seems to be reasonably good and should be used. It is the El-Emam et al. model. This could be reasonably accepted as the model was developed using the four sets of data and was evaluated using pipeline field data. Its R^2 value is the highest, 0.92, and it is ranked first. Also, the same conclusion can be drawn from the AIC results as the El-Emam et al. (2003) model is still ranked number one with an Akaike weight factor of 99.9 % and no information loss, i.e., $\Delta = 0.0$. Furthermore, since the El-Emam et al.

Table 4 Ranking of the correlations using Yoo (1974) data

Model	R^2	R^2 ranking	Δ	ω	AIC ranking
D & M	0.80	2	44.0	0.00	Poor
S & M	0.08	Poor	106.8	0.00	Poor
Tomita	0.28	Poor	153.1	0.00	Poor
Thomas	0.74	5	53.7	0.00	Poor
Clapp	0.79	3	44.5	0.00	Poor
K & K	0.21	Poor	95.1	0.00	Poor
SBN	0.39	Poor	85.2	0.00	Poor
G & S	0.11	Poor	129.5	0.00	Poor
D & E	0.94	1	0.0	1.00	1
HSW	0.78	4	48.4	0.00	Poor
El-Emam et al.	0.63	6	67.1	0.00	Poor

Table 5 Ranking of the correlations using Szilas et al. (1981) data

Model	R^2	R^2 ranking	Δ	ω	AIC ranking
D & M	0.70	7	532.2	0.00	Poor
S & M	0.84	4	487.2	0.00	Poor
Tomita	0.59	9	648.3	0.00	Poor
Thomas	0.97	3	376.9	0.00	Poor
Clapp	0.71	6	529.2	0.00	Poor
K & K	0.45	Poor	574.1	0.00	Poor
SBN	0.99	1	0.0	1.00	1
G & S	0.15	Poor	625.7	0.00	Poor
D & E	0.74	5	521.2	0.00	Poor
HSW	0.67	8	538.9	0.00	Poor
El-Emam et al.	0.99	1	245.9	0.00	Poor

Table 6 Ranking of the correlations using all data

Model	R^2	R^2 ranking	Δ	Ω	AIC ranking
D & M	0.74	7	219.321	0.000	Poor
S & M	0.86	3	99.852	0.000	Poor
Tomita	0.29	Poor	515.978	0.000	Poor
Thomas	0.82	4	147.375	0.000	Poor
Clapp	0.78	5	186.358	0.000	Poor
K & K	0.48	Poor	347.898	0.000	Poor
SBN	0.91	2	14.447	0.001	Poor
G & S	0.06	Poor	457.298	0.000	Poor
D & E	0.78	6	193.122	0.000	Poor
HSW	0.69	8	252.957	0.000	Poor
El-Emam et al.	0.92	1	0.0	0.999	1

(2003) model turned out to be the best model for a wide range of generalized Reynolds numbers and flow behavior indices, it is recommended to be used for the prediction of the friction factor.

From Table 6, we can also notice that even though the Szilas et al. model has a large R^2 value, it is a poor model from the AIC point of view. In fact with the exception of the El-Emam et al. model, all models should not be considered for the prediction.

6 Conclusions and recommendations

The present paper shows that a large number of equations exist to calculate the friction factor for pseudoplastic power law fluids. Yet, selecting the equation represents an immense challenge facing the pipeline engineer. A wrong selection may lead to an error of up to 83.4 % (El-Emam et al. 2003). Eleven equations are discussed and examined using four sets of friction factor measured data. Traditionally, R^2 along with the AIC approach are used throughout the comparative study to select the best model to predict the Fanning friction factor. Both AIC and the R^2 methods suggest that the El-Emam et al. model is reasonably good in predicting friction factors. The suggested model has the highest R^2 (0.92) as well as the highest Akaike weight factor ($\omega = 99.9$ %) with no formation loss ($\Delta = 0.0$). Moreover, this model, unlike other models, covers a wide range of both flow behavior indices and generalized Reynolds number. Nevertheless, other models showed excellent performance when compared with their original data.

The shortcomings of using R^2 are discussed where certain models can have high R^2 values, yet the Akaike weight factors are very low as an indication of their poor performance. A good example is the Szilas et al. model when examined using all the data. The advantages of using AIC over R^2 are presented which makes it a viable alternative for model selection. It employs the parsimonious principle to trade between the complexity of the model and its accuracy, not only to find a best approximating model, but also to develop statistical inference based on the data.

It is therefore recommended that the El-Emam et al. model is used to predict the Fanning friction factor employing the AIC approach rather than the conventional R^2 approach for model selection.

Finally, the authors introduce AIC to the oil and gas industry as an innovative tool for model selection. We believe this AIC can alleviate the dilemma of model selection encountered by professionals in the oil and gas industry.

References

Akaike H. Information theory as an extension of the maximum likelihood principal. Second international symposium on information theory. Budapest: Akademiai Kiado; 1973. p. 267–81.

Anderson D. Model based inference in the life sciences: a primer on evidence. New York: Springer; 2008.

Bogue D. Velocity profiles in turbulent non-newtonian pipe flow. PhD dissertation. University of Delaware, Newark; 1962.

Burnham D, Anderson D. Model selection and multimode inference, a practical information-theoretic approach. 2nd ed. New York: Springer; 2002.

Claeskens G, Hjort N. Model selection and model averaging. 2nd ed. Cambridge: Cambridge University Press; 2009.

Clapp R. Turbulent heat transfer in pseudoplastic non-newtonian fluids. Int. Developments in Heat Transfer, ASME. 1961; Part III, Sec. A, 652.

Derby R, Melson J. How to predict the friction factor for flow of bingham plastics? Chem Eng. 1981;28:59–61.

Desouky S, EL-Emam N. A generalized pipeline design correlation for pseudoplastic fluids. J Can Pet Technol. 1990;29(5):48–54.

Dodge D, Metzner A. Turbulent flow of non-newtonian systems. AIChE J. 1959;5(2):189–203.

El-Emam N, Kamel AH, Al-Shafie M, et al. New equation calculates friction factor for turbulent flow of non-Newtonian fluids. Oil Gas J. 2003;101(36):74–83.

Gao P, Zhang J. New assessment of friction factor correlations for power law fluids in turbulent pipe flow: a statistical approach. J Central South Univ Technol. 2007;14:77–81.

Garica E, Steffe J. Comparison of the friction factor equations for non-Newtonian fluids in pipe flow. J Food Process Eng. 1986;9(2):93–120.

Govier G. The flow of complex mixtures in pipes. 2nd ed. New Orleans: Society of Petroleum Engineers; 2008.

Hanks R, Dadia BH. Theoretical analysis of the turbulent flow of non-Newtonian slurries in pipes. AIChE J. 1971;17(3):554–7.

Hanks R, Ricks B. Transitional and turbulent pipe flow of pseudoplastic fluids. J Hydronaut. 1975;9:39–44.

Hartnet J, Kostic M. Turbulent friction factor correlations for power law fluids in circular and non-circular channels. Int Commun Heat Mass Trans. 1990;17:59–65.

Hawase Y, Shenoy A, Wakabayashi K. Friction and heat and mass transfer for turbulent pseudoplastic non-Newtonian fluids flowing in rough pipes. Can J Chem Eng. 1994;72:798–804.

Hemeida A. Friction factor for yieldless fluids in turbulent pipe flow. J Can Pet Technol. 1993;32(1):32–5.

Irvine TF Jr. A generalized Blasius equation for power law fluids. Chem Eng Commun. 1988;65:39–47.

Kemblowski Z, Kolodzie J. Flow resistances of non-Newtonian fluids in transitional and turbulent-flow. Int Chem Eng. 1973;13(2): 265–79.

Khaled S. A comparative study among the different methods used for calculating the pressure losses in pipelines transporting non-Newtonian oils. MSc thesis. AL-Azhar University, Egypt; 1994.

Larson R, Marx M. An introduction to mathematical statistics and its applications. 3rd ed. New Jersey: Pearson; 2007.

Mendenhall W, Sincich T. A second course in statistics: regression analysis. 6th ed. New Jersey: Pearson Education Inc; 2003.

Moody L. Friction factors for pipe flow. Trans ASME. 1944;66: 671–84.

Shaqlaih A. Model selection by an information theory approach. PhD dissertation. University of Oklahoma, Norman; 2010.

Shaqlaih A, White L, Zaman M. Resilient modulus modeling with information theory approach. Int J Geomech. 2013;13(4):384–9.

Shaver R, Merrill E. Turbulent flow of pseudoplastic polymer solutions in straight cylindrical tubes. AIChE J. 1959;5(2): 181–8.

Shenoy A. Rheology of highly filled polymers melt systems. In: Cheremisinoff NP, editor. Encyclopedia of fluid mechanics, vol. 7. Houston: Gulf Publishing; 1988. pp. 667–71.

Shenoy A, Saini D. A new velocity profile model for turbulent pipe-flow of power-law fluids. Can J Chem Eng. 1982;60(5):694–6.

Streeter V, Wylie B. Fluid mechanics. 7th ed. New York: McGraw-Hill College; 1985.

Szilas A, Bobok E, Navratil L. Determination of turbulent pressure losses of non-Newtonian oil flow in rough pipes. Rheol Acta. 1981;20(5):487–96.

Tam K, Tiu C. A general correlation for purely viscous non-Newtonian fluids flowing in ducts of arbitrary cross sections. Can J Chem Eng. 1988;66:542–9.

Thomas DG. Heat and momentum transport characteristics of non-Newtonian aqueous thorium oxide suspensions. AIChE J. 1960;6(4):631–9.

Tomita Y. A study of non-Newtonian flow in pipelines. Bull JSME. 1959;5:10–16.

Torrance B. Friction factors for turbulent non-Newtonian fluid flow in circular pipes. S Afr Mech Eng. 1963;13:89–91.

Trinh K. A boundary layer theory for turbulent transport phenomena, MSc thesis. University of Canterbury New Zealand; 1969.

Trinh K. A zonal similarity analysis of wall-bounded turbulent shear flows. Paper presented at the 7th world congress of chemical engineering; 2005.

Yoo SS. Heat transfer and friction factors for non-newtonian fluids in turbulent pipe flow. PhD dissertation. Graduate College, University of Illinois, Chicago; 1974.