Petroleum 9 (2023) 629–646

Contents lists available at ScienceDirect

Petroleum

journal homepage: www.keaipublishing.com/en/journals/petlm

An adaptive neuro-fuzzy inference system white-box model for real-time multiphase flowing bottom-hole pressure prediction in wellbores

Chibuzo Cosmas Nwanwe [a, b, *], Ugochukwu Ilozurike Duru [b]

[a] Department of Minerals and Petroleum Resources Engineering Technology, Federal Polytechnic Nekede, Owerri, P.M.B., 1036, Owerri, Nigeria
[b] Department of Petroleum Engineering, Federal University of Technology, Owerri, P.M.B., 1526, Owerri, Nigeria

ARTICLE INFO

Article history:
Received 28 June 2022
Received in revised form
23 December 2022
Accepted 31 March 2023

Keywords:
Machine learning models
Empirical correlations
Mechanistic models
Multiphase flowing bottom-hole pressure
Adaptive neuro-fuzzy inference system
White-box model

ABSTRACT

The majority of published empirical correlations and mechanistic models are unable to provide accurate flowing bottom-hole pressure (FBHP) predictions when real-time field well data are used. This is because the empirical correlations and the empirical closure correlations for the mechanistic models were developed with experimental datasets. In addition, most machine learning (ML) FBHP prediction models were constructed with real-time well data points and published without any visible mathematical equation. This makes it difficult for other readers to use these ML models since the datasets used in their development are not open-source. This study presents a white-box adaptive neuro-fuzzy inference system (ANFIS) model for real-time prediction of multiphase FBHP in wellbores. 1001 real well data points and 1001 normalized well data points were used in constructing twenty-eight different Takagi–Sugeno fuzzy inference systems (FIS) structures. The dataset was divided into two sets; 80% for training and 20% for testing. Statistical performance analysis showed that a FIS with a 0.3 range of influence and trained with a normalized dataset achieved the best FBHP prediction performance. The optimal ANFIS black-box model was then translated into the ANFIS white-box model with the Gaussian input and the linear output membership functions and the extracted tuned premise and consequence parameter sets. Trend analysis revealed that the novel ANFIS model correctly simulates the anticipated effect of input parameters on FBHP. In addition, graphical and statistical error analyses revealed that the novel ANFIS model performed better than published mechanistic models, empirical correlations, and machine learning models. New training datasets covering wider input parameter ranges should be added to the original training dataset to improve the model's range of applicability and accuracy.

1. Introduction

Multiphase flow is the simultaneous flow of a mixture of two or more phases (gas, liquid, or solid phases) [1]. Multiphase flow in circular pipes is often encountered in chemical, nuclear, petroleum, and geothermal industries [2–4]. In addition, multiphase flow is more complex than single-phase flow in predicting flowing bottom-hole pressure (FBHP) [5] because of the presence of flow patterns that cannot be easily identified [6]. The modelling of multiphase flow in pipes and its associated problems have challenged petroleum engineers for quite some time [7].

Accurate prediction of multiphase FBHP in production wells is an essential factor for the analysis and production optimization of oil and gas wells. Traditionally, FBHP data is obtained by direct measurement using downhole gauges. The exposure to mechanical failures of these downhole gauges implies that they require regular maintenance and calibration to avoid erroneous FBHP measurements [8]. Furthermore, the deployment of downhole gauges for the direct acquisition of FBHP measurements is a risky and expensive duty [9]. For these reasons, petroleum engineers have over the years developed ingenious models for accurate multiphase FBHP prediction from surface readings [9]. The models developed for predicting multiphase FBHP from surface readings include

* Corresponding author.
E-mail address: nwanwe.chibuzo@fpno.edu.ng (C.C. Nwanwe).
Peer review under responsibility of Southwest Petroleum University.

Production and Hosting by Elsevier on behalf of KeAi

https://doi.org/10.1016/j.petlm.2023.03.003

mechanistic models, empirical correlations, and most recently machine learning models. These models are referred to as tubing performance relationship (TPR) models. The operating flow rate and pressure of a production well can be established by coupling any of these TPR models with an inflow performance relationship (IPR) model such as the Vogel [10] IPR model. Operating flow rates and pressures of production wells were also analysed by Duru et al. [11] using Petroleum Expert Software to develop an integrated production model (IPM). The IPM consists of a reservoir, wellbore, and surface facility models described with MBAL (material balance equation), PROSPER (IPR and TPR) and GAP software respectively. Their proposed IPM could serve as a guide to petroleum engineers on which artificial lift system to be installed or not. Nwanwe et al. [12] developed an optimum tubing size selection model by coupling the Poettmann & Carpenter [13] TPR model computed for a wide range of tubing sizes (0.824–6.0 inch) and the Vogel [10] IPR model computed for present and future reservoir pressures. The optimum tubing was selected based on economic considerations and the present and future operating flow rates.

Oil and gas production forecast is a very important application of these TPR models. With the knowledge of nodal analysis and material balance, it is possible to forecast future oil and gas production rates and cumulative production. In the case of a solution gas drive oil reservoir, gas-liquid two-phase flow behaviour prevails in the reservoir and wellbore. In performing production forecast, a series of average reservoir pressures, $\overline{p}$ are assumed between the bubble point and abandonment pressures. The material balance model presented in Craft & Hawkins [14] is then used to determine the incremental cumulative production and cumulative production for each interval of the assumed $\overline{p}$. Similarly, Nodal analysis is used in the production rate determination for each interval of assumed $\overline{p}$ by coupling a TPR and IPRs determined with the assumed $\overline{p}$. The production time for each interval of $\overline{p}$ is given as the incremental production divided by the production rate while the cumulatve production time is the sum of the production time. Future production forecast can also be performed by coupling of a capacitance resistance model with nodal analysis.

The capacitance resistance model (CRM) is a data-driven nonlinear model for the estimation of interwell connectivity (IWC) based on material balance equation between the well injection rate as input and well production rate as output within a specific control volume [15,16]. Yousef [17] proposed a CRM to infer IWC and storage from injection and production rates by coupling the material balance and the linear productivity equations. To study the IWCs, the authors proposed the balanced capacitance model (BCM) approach if the waterflood is balanced and unbalanced capacitance model (UCM) approach if the waterflood is not balanced. Sayarpour et al. [18] derived analytical solutions for the CRM based on the superposition in time and presented the solutions for three different control volumes based on tank (CRMT), producer (CRMP), and injector-producer (CRMIP) control volumes. Weber et al. [19] coupled their developed CRM with an optimizer and tested the coupled model with field data. Nguyen et al. [20] derived an integrated capacitance-resistance model (ICRM) for primary recovery from a material balance equation assuming no aquifer, black oil, and no initial gas cap. Kim et al. and Nguyen et al. [15,20] proposed an ICRM for secondary recovery from the differential equation of the producer based representation of CRM (CRMP). Yousefi et al. [16] developed a modified CRM as an analytical solution in gas-oil systems by coupling a modified material balance equation and a modified productivity equation. Wang et al. [21] presented an improved CRMIP developed by coupling the effect of active aquifer support and Stochastic Simplex Approximate Gradient optimization algorithm to the traditional CRMIP for accurate IWC determination. Real-time prediction of FBHP can improve the use of CRMs for primary, secondary, and tertiary oil recoveries since the works of Holanda, et al. and Sayarpour et al.[18,22]; and Wang, et al. [23] emphasize its importance.

The empirical correlations developed over the years for multiphase FBHP prediction are categorized as category A, B, and C correlations respectively [24]. Flow patterns and slip conditions were neglected in the development of Category A correlations [13,25,26]. In developing Category B correlations, slip was considered and flow pattern neglected [27,28]. Category C correlations considered flow pattern and slip conditions [29–31]. These empirical correlations were developed by establishing mathematical equations with the aid of experimental data [32]. The range of applicability of the empirical correlations is limited to the experimental data range used in their development. In addition, the graphical nature of the empirical correlations makes their coding in a computer program difficult.

Mechanistic models for multiphase FBHP prediction have been classified by Gomez et al. [33] as pipeline, wellbore, and unified mechanistic models. Pipeline mechanistic models are applicable for horizontal flow conditions [34,35], wellbore mechanistic models for vertical to slightly inclined flow conditions [2–4,36–38] and unified mechanistic model for all angles of inclination flow conditions [33,39–41]. The development of mechanistic models was achieved using physical laws and then closed using closure empirical correlations. Unlike empirical correlations, mechanistic models don't have any graphical correlations and hence coding in a computer program is easy. The aim of developing mechanistic models was that they exceed the performance of current empirical correlations [42]. However, statistical error analysis indicates that the published empirical correlations outperformed mechanistic models [43]. It has been recently confirmed that the poor performance of mechanistic models was due to the severe underprediction of the pressure drop by the mechanistic slug flow and churn flow models [44].

It is evident from statistical error analysis conducted by Takacs [43] that mechanistic models and empirical correlations produced unsatisfactory FBHP predictions. This has motivated researchers to develop models for predicting FBHP based on machine learning (ML). ML models are classified as genetic algorithms (GA), artificial neural networks (ANN), expert systems (ES), and fuzzy logic controllers (FLC) [45]. ANN are the most common ML models used to predict FBHP in the petroleum industry. The findings of several machine learning research [1,42,46–48] indicate that ANN provides better predictions of FBHP than empirical correlations and mechanistic models.

Adaptive neuro-fuzzy inference System (ANFIS), a combination of ANN and fuzzy logic controller (FLC) [49], is another ML model that has gained popularity in the petroleum industry in solving complex and difficult problems. ANFIS has been used in the petroleum industry for lost circulation prediction [50], natural fractures and hydraulic fractures interaction prediction [51], prediction of porosity and permeability of oil reservoirs [52], prediction of nanofluid convective flow pressure in a heated pipe [53], differential pressure prediction for downhole inflow control devices [54], prediction of reservoir bubble point pressure [55] and FBHP prediction [8,56]. ANFIS and several of its hybrid versions have also been used in other fields of study for the prediction of soil moisture [57], dew point temperature [58], daily reference evapotranspiration [59], monthly actual evapotranspiration [60], and cation exchange capacity [61]. These hybrid ANFIS models are classical ANFIS model coupled with metaheuristic algorithms such as whale optimization algorithm (ANFIS-WOA), the krill herd algorithm (ANFIS-KHA), and the firefly algorithm (ANFIS-FA), the bee colony optimization (ANFIS-BCO), the dragonfly algorithm (ANFIS-DFA), the shuffled frog-leaping algorithm (ANFIS-SFLA) and invasive

weed optimization (ANFIS-IWO), the grey wolf optimization (ANFIS-GWO), differential evolution (ANFIS-DE), and particle swarm optimization (ANFIS-PSO). Literature review conducted in this study reveals that these hybrid ANFIS models perform better than the classical ANFIS models. These successful applications of ANFIS in solving complex problems in petroleum engineering and other fields of study are indicative of the potential of ANFIS in predicting FBHP. However, these ANFIS models were developed and presented as black-box models without any visible mathematical equation making their use by other authors difficult since the data used in their development are not open-source.

Detailied literature review carried out in this study identified some issues with published correlations and models for FBHP prediction in wellbores. First, majority of mechanistic models and empirical correlations predict unsatisfactory FBHP when real-time well datasets are used. This is because the empirical correlations and the empirical closure correlations of the mechanistic models were developed with experimental datasets. Second, real-time well data points are used in the construction of the majority of ML FBHP prediction models. However, these ML models are published without any visible mathematical equation aside from the ANN-based models developed by Nwanwe et al. and Tariq et al. [47,62]. This makes the use of these ML models difficult since the data employed in their development are not open-source. Lastly, most ML models are ANN-based and to the best of our knowledge, the [8,56] model is the only ANFIS-based FBHP prediction model reported in the literature. The [8,56] model was constructed using 795 real-time data points and presented without any visible mathematical equation. The author only described how he trained the dataset to obtain the best black-box model with a range of influence of 0.6. This makes it difficult to use his ANFIS-based model for FBHP determination on new dataset.

An ANFIS-based white-box FBHP prediction model is presented. 1001 real-time normalized well data points [56,63] were employed in the black-box model construction. The dataset was divided into two sets; 80% for training and 20% for testing. The dataset division ensured that the testing dataset parameters ranges were within the training dataset parameter ranges. Feature selection conducted showed that seven out of the eight input parameters are strongly correlated to the FBHP. However, the authors decided to develop the model with all eight input parameters because field experience shows that all these parameters affect FBHP. In addition, feature scaling performed normalized the ranges of the input and output parameters between 0 and +1. Subtractive clustering was used in creating the Takagi–Sugeno fuzzy inference system (FIS) in this study. Several Takagi–Sugeno FIS were created, each defined by a range of influence (RI) ranging from 0.3 to 0.95 and type of dataset (normalized or real dataset). After training each of these FIS, the best FIS with a RI of 0.3 and trained with normalized was selected. The tuned premise and consequence parameter sets of the selected FIS black-box model were extracted and used to implement the ANFIS white-box model. The developed ANFIS visible mathematical model was also implemented in microsoft office excel and included as a supplementary file. Users will therefore be able to apply the current ANFIS visible mathematical model in the field for predicting FBHP in real-time from real-time well data. Performance evaluation of the developed ANFIS visible mathematical model was done by performing trend analysis and comparison against three mechanistic models, three empirical correlations, one ANN-based model and one ANFIS-based model.

2. Literature review

Many researchers have for many years developed ML models for determination of multiphase FBHP in wellbores. This section reviews some of the ML models that have been built for FBHP prediction. This section also reviews the application of ANFIS in various areas.

Ahmadi and Chen [64] developed a model based on ANN for predicting FBHP. The authors improved the predictive performance of the ANN-based model by optimizing the weights and biases with the aid of meta-heuristic optimization algorithms. The results of their study showed that particle swarm optimization (PSO) and hybrid genetic algorithm (HGA) performed best. Furthermore, the findings suggested that hybrid models might be applied in commercial production engineering software.

Akinsete and Adesiji [65] collected and processed 15,635 data points from a Norwegian field into 2447 data points. The 2447 data points were then used in constructing a model based on ANN for predicting FBHP. The ANN model was trained and tested using 1713 and 734 data points respectively. The results of their study showed that their ANN model performed better than support vector, random forest, decision trees, and mechanistic models.

Al-Shammari [8,56] collected and reduced 1207 well data points from Middle East oil fields to 795 data points by processing with empirical correlations to remove outliers. The 795 well data points were then used to construct an ANFIS model that employs subtractive clustering with a 0.6 cluster radius in predicting FBHP. The author used 596 and 199 well data points respectively for training and testing of the model. A comparative study showed that the ANFIS model performed better than existing empirical correlations.

El-Kadi et al. [66] used a database of 7581 data points in constructing a model based on ANN for two-phase pressure drop prediction in oil wells. Independent results were generated by the authors by considering the effect of the number of hidden layers and the number of neurons in each hidden layer. A comparative study revealed that their ANN model accurately predicts pressure drop than published mechanistic models and empirical correlations.

Jahanandish et al. [42] collected 413 Iranian oil well data points and used the data points to construct a model based on ANN for predicting multiphase FBHP. The dataset was partitioned in a ratio of 4:1:1 for the training, testing, and validation datasets respectively. The results showed that the developed ANN model performed with higher accuracy than published mechanistic models and empirical correlations.

Kanin et al. [1,67] proposed a novel method for the computation of multiphase pressure gradient that involves three different machine learning models. In the developed model, the first surrogate model is used for liquid holdup prediction, the second for flow pattern prediction, and the third for pressure gradient prediction. Experimental data from open source was used to evaluate their ML model and was shown to outperform published mechanistic models and empirical correlations. It was also shown that their ML model permitted a wider input data range.

Li et al. [46] proposed a procedure for FBHP computation that couples the FBHP computation procedures of ANN models and empirical correlations to improve the prediction accuracy and increase the range of prediction. Their combined procedure provided better prediction results than the existing empirical correlations.

Osman et al. [45] developed a model based on ANN for predicting FBHP using 206 data points collected from Middle East fields. The authors divided the dataset into training, validation, and testing datasets in a ratio of 3:1:1. The results revealed that their ANN model outperformed current mechanistic models and empirical correlations. In addition, the trend analysis conducted showed that their model provided the physical parameters' expected effects on the pressure drop.

Sami and Ibrahim [48] developed ML models for predicting multiphase FBHP by employing ANN, Random Forest (RF), and K-

Nearest Neighbours (KNN). A comparison of the three ML models used in their study showed that FBHP is predicted with higher accuracy by ANN than RF and KNN. Their study also showed that artificial intelligence (AI) has the potential of predicting complex multiphase petroleum production parameters.

Seong et al. [68] used 350 experimental data points to construct a deep neural network (DNN) model for the evaluation of multiphase horizontal flow in pipes. The authors used 279, 3, and 40 experimental data points for training, validation, and forecasting respectively. Coefficients of determination of 0.89 and 0.98 were reported for liquid holdup and pressure gradient predictions respectively which is indicative of the model's reliability.

Tariq et al. [62] used the Ayoub [63] 206 real-time well data points to develop a black-box model based on ANN for predicting FBHP in vertical wells. The trained black-box model's weights and biases were extracted and used to implement the white-box model. The authors also used particle swarm optimization (PSO) in optimizing the weights and biases of the developed ANN white-box model. A comparative study showed that the PSO-ANN FBHP prediction model outperformed the conventional ANN FBHP prediction model. Furthermore, group trend analysis revealed that the PSO-ANN FBHP prediction model captures the physics behind the problem accurately.

Moazenzadeh et al. [57] evaluated the performance of a classical ANFIS model and three ANFIS models coupled respectively with the whale optimization algorithm (ANFIS-WOA), the krill herd algorithm (ANFIS-KHA), and the firefly algorithm (ANFIS-FA) in predicting soil moisture (SM). The data consist of SM content and soil temperature measured at a depth of 20 cm at different locations in Turkey for a total of 490 days as well as meteorological data measured at Istanbul Bolge station. The result of their study revealed that all three hybrid ANFIS models performed better than the classical ANFIS with the ANFIS-WOA achieving the best performance.

Mehdizadeh et al. [58] developed a classical ANFIS model and two ANFIS models coupled respectively with the bee colony optimization (ANFIS-BCO), and the dragonfly algorithm (ANFIS-DFA) for the prediction of dew point temperature. The data used for the study was collected from the Rasht and Urmia meteorological stations in Iran. A comparative study revealed that the hybrid models (ANFIS-BCO and ANFIS-DFA) outperformed the classical ANFIS model with the ANFIS-DFA performing best for the considered stations.

Mehdizadeh et al. [59] developed two ANFIS hybrid models in which ANFIS is coupled respectively with the shuffled frog-leaping algorithm (ANFIS-SFLA) and invasive weed optimization (ANFIS-IWO) in the prediction of daily reference evapotranspiration. The study data was collected from the Tabriz and Shiraz meteorological stations in Iran. Comparative evaluation conducted showed that the two ANFIS hybrid models outperformed the classical ANFIS model and existing empirical models, with the ANFIS-SFLA achieving the best performance.

Hadadi et al. [60] evaluated the performance of classical ANFIS model and two ANFIS models coupled respectively with the shuffled frog-leaping algorithm (ANFIS-SFLA) and the grey wolf optimization (ANFIS-GWO) in the prediction of monthly actual evapotranspiration. The data used for their study were collected in Neishaboor watershed, Iran and consist of meteorological data and remotely sensed data. The result revealed that the ANFIS-SFLA and ANFIS-GWO outperformed the classical ANFIS model with ANFIS-GWO exhibiting the overall best performance.

Emamgholizadeh et al. [61] developed classical ANFIS model and two ANFIS models coupled respectively with differential evolution (ANFIS-DE), and particle swarm optimization (ANFIS-PSO) for the prediction of cation exchange capacity (CEC). The data employed in their study consist of 250 topsoil samples collected in the Gilan province of Iran. The authors divided the dataset into two sets; 75% for training and 25% for testing. The study result revealed the superior performance of the two hybrid ANFIS models over the classical ANFIS model in predicting CEC.

3. Adaptive neuro-fuzzy inference system (ANFIS)

ANFIS, a machine learning technique developed in the early 90s [69], is a type of ANN based on the Takagi–Sugeno fuzzy inference system. ANFIS allows users to benefit from the advantages of fuzzy logic and neural network concepts because it integrates both concepts in a single framework. The ANFIS structure consists of five layers [49].

(1) The input values go through the first layer (fuzzification layer), which decides which membership functions apply to them. The basic types of membership functions include Gaussian, trapezoidal, triangular, and bell-shaped membership functions. The Gaussian membership function (gaussmf) was used in the development of the current model and hence will be the focus here. The premise parameter set, $\{\sigma, c\}$ is used to calculate the membership degrees of each function as given in Eq. (1). The gaussmf has a smooth curve and uses the parameter σ to locate the centre of the curve and parameter c for determining the width of the curve [70].
(2) The second layer (rule layer) multiplies the membership values calculated in the first layer to generate the firing strengths for the rules, as shown in Eq. (2). Note that this is for a scenario with only two inputs.
(3) The third layer (normalization layer) normalizes the computed firing strengths by dividing each value of firing strength by the total firing strength as given in Eq. (3).
(4) The normalized data and the consequence parameter set {p, q, r} are fed into the fourth layer (defuzzification layer) which returns the defuzzified values as given in Eq. (4).
(5) The defuzzified values are then transferred to the fifth layer (summation layer), which sums them and returns the final output as given in Eq. (5).

$$O_i^1 = \mu_{A_i}(x) = gaussmf(x;\sigma,c) = \exp\left[-0.5\left(\frac{x-c}{\sigma}\right)^2\right] \tag{1}$$

$$O_i^2 = w_i = \mu_{A_i}(x).\mu_{B_i}(x), i = 1,2 \tag{2}$$

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2 + w_3 + w_4},\ i \in \{1,2,3,4\} \tag{3}$$

$$O_i^4 = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{4}$$

$$O_i^5 = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{5}$$

where $\{\sigma, c\}$ is the premise parameter set; $\{A_i, B_i\}$ is the linguistic label to node i, x is the input to node i, w_i is the firing strength, $\overline{w_i}$ is the normalized firing strength from layer 3, $\{p, q, r\}$ is the consequence parameter set.

The approaches proposed over the years for the training of ANFIS can be classified as derivative based, heuristic based, and derivative-heuristic based hybrid approaches [49]. As already mentioned above, ANFIS consist of premise and consequence parameter sets. Literature review reveals that there are two

approaches for the training of premise and consequence parameter sets. In the first approach, premise and consequence parameter sets are trained with the same training optimization algorithm. In the second approach, the premise parameter set is trained with an optimization algorithm different from that used in training the premise parameter set.

The derivative based learning algorithm can further be divided into standard and hybrid derivative training approaches. In standard derivative training approach, the premise and consequence parameter sets are trained with the same training optimization algorithm, e.g. backpropagation (BP) [71] or Levenberg-Marquardt (LM) [72] training optimization algorithm. In the derivative based hybrid training approach, the premise set is trained with an optimization algorithm different from that used in determining the consequence parameter set. The most widely used version of the derivative based hybrid algorithm was originally proposed by Jang [69] and is commonly referred to as hybrid learning (HL). In HL, the premise parameter set is trained using backpropagation gradient descent (BP/GD) while the consequence parameter set is determined with least square estimation (LSE). In other versions of derivative based hybrid algorithms, the premise parameter set is trained with BP/GD while the consequence parameter set is estimated with either recursive least square estimation (RLSE) [73], LM [74], Kalman filter (KF) [75], or Extended Kalman filter (EKF) [76]. The drawback of these derivative based hybrid algorithms stems from their use of BP/GD, which is prone to getting stuck in the local minima, to train the premise parameter set.

The heuristic based learning algorithms, classified as standard and variant, were developed to overcome the drawback of derivative based training algorithms. In fact, the trend has shifted from derivative based algorithms to heuristic based algorithms owing to the flexibility and global search ability. In heuristic based learning algorithms, the premise and consequence parameter sets are trained with the same algorithm. Common examples of standard heuristic based learning algorithm used for training the ANFIS premise and consequence parameter sets include particle swarm optimization (PSO) [77], genetic algorithm (GA) [78], artificial bee colony (ABC) [79], and firefly algorithm (FA) [80]. Accelerated particle swarm optimization (APSO) [81], Real-coded genetic algorithm (RCGA) [82], and adaptive and hybrid artificial bee colony algorithm (aABC) [83].

In heuristic-derivative based hybrid approach, the premise parameter set is trained with the heuristic algorithm while the consequence parameter set is trained with derivative algorithm. The commonly used heuristic algorithms used in hybrid training algorithms include GA and PSO. LSE and BP/GD are the most common derivative based algorithms used in hybrid training algorithms. Recently, new versions of hybrid algorithms have been reported that couples classic ANFIS with metaheuristic algorithms such as whale optimization algorithm (WOA), krill heard algorithm (KHA), shuffled frog leaping algorithm (SFLA), grey wolf optimization (GWO), invasive weed optimization (IWO) [57–61]. Heuristic-derivative based hybrid approaches have been very successful because they employ the advantages of heuristic and derivative based algorithms [49].

4. Methodology

4.1. Data collection and description

The collection of representative dataset is the initial and key stage in machine learning model development. 1001 open-sourced well data points were employed in constructing this study's ANFIS visible mathematical model for predicting FBHP. 795 and 206 well data points were collected from the thesis of Al-Shammari [56] and Ayoub [63] respectively.

1207 well data points were gathered by Al-Shammari [56] from Middle East fields and preprocessed to exclude inaccurate data points. This was achieved by using existing empirical correlations and excluding data points with an arithmetic average error greater than 15%. This resulted in the reduction of the data points to 795. The [56] dataset consists of eight independent input parameters and one dependent output parameter. The independent input parameters include well bottom-hole temperature (WBHT), internal diameter of tubing (ID), oil API gravity (API), well perforation depth (WPD), gas flow rate (GFR), oil flow rate (OFR), water flow rate (WFR), and wellhead pressure (WHP) [48]. The dependent output parameter is flowing bottom-hole pressure (FBHP).

386 well data points were gathered by Ayoub[63] from Middle East fields. The authors then employed empirical correlations and mechanistic models to exclude unrepresentative data points. This resulted in the reduction of the data points to 206. The Ayoub [63] well dataset is made up of nine independent input parameters and one dependent output parameter. The independent input parameters include wellhead temperature (WHT) and the [56] dataset's eight independent input parameters. FBHP is the output production parameter.

In the present study, WHT was removed from the Ayoub [63] dataset and the resulting dataset was combined with the [56] dataset. We divided the combined dataset into two sets; 798 data points for training and 203 for testing. Tables 1 and 2 show the statistical analysis of the training and testing datasets respectively. Notice that the parameter ranges of the testing dataset are within the training dataset parameter ranges.

4.2. Feature selection

Feature selection is the process of extracting the significant features for use in predictive model creation. The objective of feature selection is to improve the prediction performance of the model while reducing the modelling cost. Feature selection methods can be subdivided into supervised and unsupervised methods. Unsupervised feature selection methods are used for the removal of redundant input variables by ignoring the output variable. Supervised feature selection methods are used for the removal of irrelevant variables with the aid of the output variable. The classes of supervised feature selection methods include filter, wrapper, and intrinsic feature selection methods [84].

The training dataset as summarized in Table 1 contain eight independent input production parameters. Using too many input parameters in developing the model will require large system memory which can slow the model development and training. There is therefore the need to perform feature selection to exclude the insignificant independent input parameters so that the model can run efficiently.

In this study, the first step in feature selection is to choose the feature selection method. The Pearson's correlation coefficient is chosen because the inputs and output of the training dataset are numerical. Data analysis toolpak in excel was used to plot the Pearson's correlation matrix of all variables in the training dataset as shown in Table 3.

To perform the feature selection, we defined a threshold correlation coefficient (TCC) between the independent and dependent parameters. This TCC was determined with the aid of Eq. (6) and the table of critical values for Pearson's correlation coefficient (Table 4) at a significance level of 5% ($\alpha = 0.05$) [85]. For the removal of irrelevant parameters, the following criterion is used. If the CC between any of the independent parameters and the dependent parameter is less than the defined TCC, then the said independent parameter is irrelevant and is thus removed.

Table 1
Training dataset statistical analysis (798 data points).

S/N	Parameters	Units	Min	Max	Average	STD
1	Wellhead pressure (WHP)	psia	80.00	1550.00	409.51	241.93
2	Water flow rate (WFR)	stb/d	0.00	11395.00	2300.56	2413.62
3	Oil flow rate (OFR)	stb/d	176.00	19618.00	5615.47	4013.53
4	Gas flow rate (GFR)	Mscf/d	9.00	17859.00	2843.64	2556.03
5	Well perforation depth (WPD)	feet	4243.00	8620.00	6328.57	527.27
6	Oil API gravity (API)	°API	25.40	47.50	33.85	3.09
7	Internal diameter of tubing (ID)	inch	1.995	6.276	3.942	0.559
8	Well bottom-hole temperature (WBHT)	°F	157.00	233.00	209.11	18.89
9	Flowing bottom-hole pressure (FBHP)	psia	1198.00	3698.00	2489.75	371.93

Table 2
Testing dataset statistical (203 data points).

S/N	Parameters	Units	Min	Max	Average	STD
1	Wellhead pressure (WHP)	psia	95.00	1350.00	375.81	231.79
2	Water flow rate (WFR)	stb/d	0.00	10500.00	2371.86	2412.06
3	Oil flow rate (OFR)	stb/d	469.00	17243.00	5576.73	3908.91
4	Gas flow rate (GFR)	Mscf/d	75.20	12586.00	2858.66	2505.96
5	Well perforation depth (WPD)	feet	4550.00	7151.00	6353.78	514.86
6	Oil API gravity (API)	°API	27.60	42.50	33.82	2.39
7	Internal diameter of tubing (ID)	inch	1.995	6.276	3.880	0.448
8	Well bottom-hole temperature (WBHT)	°F	160.00	225.00	208.01	15.15
9	Flowing bottom-hole pressure (FBHP)	psia	1227.00	3325.00	2410.62	362.67

$$df = n - 2 \tag{6}$$

where df is the degree of freedom and n is the number of data points.

For the current study, the number of data points, n in the training dataset is equal to 798 which gives a degree of freedom, df of 796 which is approximately equal to 800. From Table 4, the critical value at $df = 800$ and $\alpha = 0.05$ is 0.069234. This critical value of 0.069234 is the TCC between the independent and dependent parameters.

As shown in Table 3, all the input parameters with exception of gas flow rate (GFR) exhibited an absolute value of CC with FBHP greater than the defined TCC of 0.069234. This implies that GFR is an irrelevant parameter and is thus removed. This shows that seven out of the eight input parameters are correlated to the FBHP. However, all eight input parameters will be included in the model development since all these parameters are known to affect FBHP in reality.

4.3. Feature scaling

Feature scaling is a data preprocessing method used in machine learning to normalize features in a dataset to be in the same range. This will ensure that no feature dominates other features during the training of the dataset. Feature scaling is recommended when the training dataset contains input parameters of different scales and units. Table 1 clearly shows that the training dataset contains input parameters of different scales and units, hence, feature selection is recommended. Feature scaling methods commonly used include scaling to unit length, standardization, mean normalization and min-max normalization. For the present study, we adopted min-max normalization given in Eq. (7) and used it to normalize the eight independent input parameters and the dependent output parameter ranges to lie between 0 and +1. The authors of this study also found that the model developed with a normalized dataset outperformed the model developed with the actual dataset ranges.

$$X_{n(0:1)} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7}$$

where $X_{n(0:1)}$ is the normalized parameter value, X is the original parameter value, X_{min} is the parameter minimum value given in column 4 of Table 1 and X_{max} is the parameter maximum value given in column 5 of Table 1.

Table 3
Pearson's correlation matrix of all production parameters in the training dataset.

	WHP	WFR	OFR	GFR	WPD	API	ID	WBHT	FBHP
WHP	1.00								
WFR	−0.23	1.00							
OFR	−0.07	−0.16	1.00						
GFR	0.06	0.01	0.83	1.00					
WPD	0.02	0.31	0.14	0.23	1.00				
API	0.41	0.22	0.21	0.44	0.37	1.00			
ID	0.04	0.12	0.26	0.19	0.00	0.00	1.00		
WBHT	0.11	0.32	0.22	0.31	0.50	0.67	0.03	1.00	
FBHP	0.51	0.38	−0.11	0.02	0.50	0.43	−0.08	0.34	1.00

4.4. ANFIS black-box model development

The ANFIS model for FBHP prediction was built and optimized using MATLAB in this study. As already mentioned, a total of 1001 data points collected from open-sources were used in developing the present ANFIS model. The dataset was divided into two sets; 798 data points for training (80%) and 203 (20%) for testing. This division ensured that the data ranges of the testing dataset were within that of the training dataset. The model was developed with the eight input parameters.

Subtractive clustering was used to generate the Takagi–Sugeno fuzzy inference system (FIS). To obtain an optimal Takagi–Sugeno FIS, a total of twenty-eight Takagi–Sugeno FIS as shown in Table 5 were created, each defined by a range of influence (RI) ranging from 0.3 to 0.95 (column 2) and type of training dataset (column 3). Each of these FIS was optimized with a hybrid optimization method

Table 4
Table of critical values for Pearson's correlation coefficient.

	α					
df	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.694390	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.519357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.060207	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.575000	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.582270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455310	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.598270	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373866	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119
35	0.215598	0.274611	0.324573	0.380976	0.418211	0.518898
40	0.201796	0.257278	0.304396	0.357787	0.393174	0.489570
45	0.190345	0.242859	0.287563	0.338367	0.372142	0.464673
50	0.180644	0.230620	0.273243	0.321796	0.354153	0.443201
60	0.164997	0.210832	0.250035	0.294846	0.324818	0.407865
70	0.152818	0.195394	0.231883	0.273695	0.301734	0.379799
80	0.142990	0.182916	0.217185	0.256525	0.282958	0.356816
90	0.134844	0.172558	0.204968	0.242227	0.267298	0.337549
100	0.127947	0.163782	0.194604	0.230079	0.253979	0.321095
125	0.114477	0.146617	0.174308	0.206245	0.227807	0.288602
150	0.104525	0.133919	0.159273	0.188552	0.208349	0.264316
175	0.096787	0.124036	0.147558	0.174749	0.193153	0.245280
200	0.090546	0.116060	0.138098	0.163592	0.180860	0.229840
250	0.081000	0.103852	0.123607	0.146483	0.161994	0.206079
300	0.073951	0.094831	0.112891	0.133819	0.148019	0.188431
350	0.068470	0.087814	0.104551	0.123957	0.137131	0.174657
400	0.060391	0.082155	0.097824	0.115997	0.128339	0.163520
450	0.057294	0.077466	0.092248	0.109397	0.121046	0.154273
500	0.057294	0.073497	0.087528	0.103808	0.114870	0.146436
600	0.052305	0.067103	0.079920	0.094798	0.104911	0.133787
700	0.048427	0.062132	0.074004	0.087789	0.097161	0.123935
800	0.045301	0.058123	0.069234	0.082135	0.090909	0.115981
900	0.042711	0.054802	0.065281	0.077450	0.085727	0.109385
1000	0.040520	0.051993	0.061935	0.073484	0.081340	0.103800
1500	0.033309	0.042458	0.050582	0.060022	0.066445	0.084822
2000	0.028654	0.036772	0.043811	0.051990	0.057557	0.073488
3000	0.023397	0.030027	0.035775	0.042457	0.047006	0.060027
4000	0.020262	0.026005	0.030984	0.036773	0.040713	0.051996
5000	0.018123	0.023260	0.027714	0.032892	0.036417	0.046512

because of the minimum error produced as compared to the backpropagation optimization method. The hybrid optimization method used in this study is derivative based hybrid optimation algorithm reffered to as hybrid learning (HL). In HL, the premise parameter set is trained using backpropagation gradient descent (BP/GD) while the consequence parameter set is determined with least square estimation (LSE). In other versions of derivative based hybrid algorithms, the premise parameter set is trained with BP/GD while the consequence parameter set is estimated with either recursive least square estimation (RLSE) [73], Levenberg-Marquardt (LM) [74], Kalman filter (KF) [75], or Extended Kalman filter (EKF) [76]. It has been reported that HL used in the present study provides better performance than most of the derivative based training approaches [49].

At the end of the training of each of the twenty-eight FIS, the 1001 normalized inputs are used in the FIS to compute the normalized FBHP. Eq. (8) was then used to de-normalize the normalized FBHP to obtain the predicted FBHP.

$$FBHP = FBHP_{n(0:1)}(FBHP_{max} - FBHP_{min}) + FBHP_{min} \tag{8}$$

where $FBHP_{n(0:1)}$ is the normalized predicted FBHP, $FBHP$ is the actual predicted FBHP, $FBHP_{max}$ is the maximum FBHP (3698 psia) and $FBHP_{min}$ is the minimum FBHP (1198 psia).

Three statistical indicators (relative performance factor, root mean square error and correlation coefficient) were used in this study for the selection of the optimum FIS. The relative performance factor, F_{rp} is given in Eq. (9) and is recommended for comparison between a group of models [2] with a value of zero and six for the best-performing and worst-performing models respectively. The root mean square error, E_{rms} is given in Eq. (10) with the minimum and maximum values representing the best-performing and worst-performing models respectively. The correlation coefficient is given in Eq. (11) with the minimum and maximum values representing worst-performing and best-performing models respectively.

$$F_{rp} = \frac{|E_1| - |E_1|}{|E_{1max}| - |E_{1min}|} + \frac{E_2 - E_{2min}}{E_{2max} - E_{2min}} + \frac{E_3 - E_{3min}}{E_{3max} - E_{3min}} + \frac{|E_4| - |E_{4min}|}{|E_{1max}| - |E_{4min}|} + \frac{E_5 - E_{5min}}{E_{5max} - E_{5min}} + \frac{E_6 - E_{6min}}{E_{6max} - E_6} \tag{9}$$

$$E_{rms} = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n} {e_{ri}}^2\right)} \tag{10}$$

$$R = \sqrt{1 - \left(\frac{\sum_{i=1}^{n} e_i}{\sum_{i=1}^{n} \overline{e}_{imeas}}\right)} \tag{11}$$

$$e_{ri} = \frac{\Delta p_{i_{pred}} - \Delta p_{i_{meas}}}{\Delta p_{i_{meas}}} \tag{12}$$

$$e_i = \Delta p_{i_{pred}} - \Delta p_{i_{meas}} \tag{13}$$

$$E_1 = \frac{1}{n}\left(\sum_{i=1}^{n} e_{ri}\right) \tag{14}$$

$$E_2 = \frac{1}{n}\left(\sum_{i=1}^{n} |e_{ri}|\right) \tag{15}$$

$$E_3 = \sum_{i=1}^{n} \sqrt{\frac{(e_{ri} - E_1)^2}{n-1}} \tag{16}$$

$$E_4 = \frac{1}{n}\left(\sum_{i=1}^{n} e_i\right) \tag{17}$$

Table 5
28 different Takagi–Sugeno FIS structures and their performances.

S/N	Range of influence	Training data type	E_1 (%)	E_2 (%)	E_3 (%)	E_4 (psia)	E_5 (psia)	E_6 (psia)	F_{rp}	E_{rms}	R
1	0.30	Real	0.755	4.662	23.670	7.358	93.058	2939.592	0.424	0.066	0.899
2	0.30	Normalized	0.729	**4.579**	22.845	7.322	**91.287**	**2885.593**	**0.245**	**0.065**	**0.902**
3	0.35	Real	0.794	4.825	24.891	7.905	96.699	3059.227	0.878	0.067	0.894
4	0.35	Normalized	0.710	4.721	22.249	7.009	94.605	2998.340	0.288	0.066	0.899
5	0.40	Real	0.828	5.131	25.957	7.535	102.336	3227.437	1.192	0.070	0.886
6	0.40	Normalized	**0.699**	4.906	**21.897**	**6.723**	97.881	3093.118	0.366	0.068	0.893
7	0.45	Real	0.940	5.546	29.473	8.903	111.383	3504.610	2.359	0.075	0.865
8	0.45	Normalized	0.853	5.029	26.728	8.990	101.120	3179.701	1.561	0.070	0.884
9	0.50	Real	0.917	5.468	28.743	9.080	109.403	3440.239	2.236	0.074	0.869
10	0.50	Normalized	0.894	5.397	28.017	8.992	108.036	3396.833	2.070	0.073	0.874
11	0.55	Real	1.066	5.801	33.398	9.704	115.459	3626.015	3.167	0.079	0.856
12	0.55	Normalized	0.869	5.094	27.231	8.456	102.193	3220.350	1.532	0.069	0.887
13	0.60	Real	1.090	5.781	34.160	9.919	115.946	3649.785	3.309	0.077	0.851
14	0.60	Normalized	0.872	5.079	27.340	8.810	102.188	3220.219	1.631	0.068	0.887
15	0.65	Real	1.140	5.926	35.727	10.474	118.861	3735.986	3.756	0.079	0.845
16	0.65	Normalized	0.940	5.077	29.469	9.605	101.191	3191.107	1.998	0.072	0.884
17	0.70	Real	1.107	5.882	34.678	10.116	117.620	3702.978	3.508	0.079	0.851
18	0.70	Normalized	0.923	5.065	28.934	9.334	100.907	3178.196	1.861	0.072	0.885
19	0.75	Real	1.115	5.857	34.934	10.469	117.524	3695.377	3.609	0.078	0.852
20	0.75	Normalized	0.894	5.018	28.016	9.505	100.850	3176.266	1.804	0.068	0.886
21	0.80	Real	1.407	7.274	44.096	10.301	145.510	4580.335	5.954	0.096	0.770
22	0.80	Normalized	0.956	5.478	29.974	8.652	109.489	3459.954	2.251	0.073	0.870
23	0.85	Real	1.404	7.258	44.013	10.387	145.171	4569.294	5.951	0.096	0.772
24	0.85	Normalized	0.957	5.475	29.991	8.701	109.452	3457.694	2.262	0.073	0.870
25	0.90	Real	1.388	7.255	43.482	10.124	145.219	4568.888	5.833	0.096	0.771
26	0.90	Normalized	0.950	5.464	29.771	8.584	109.058	3448.434	2.194	0.073	0.871
27	0.95	Real	1.180	7.143	36.978	7.461	142.404	4496.656	4.401	0.095	0.781
28	0.95	Normalized	0.957	5.470	30.005	8.809	109.395	3457.103	2.289	0.073	0.870

$$E_5 = \frac{1}{n}\left(\sum_{i=1}^{n}|e_i|\right) \tag{18}$$

$$E_6 = \sum_{i=1}^{n}\sqrt{\frac{(e_i - E_4)^2}{n-1}} \tag{19}$$

$$\overline{e}_{imeas} = \Delta p_{i_{meas}} - \overline{\Delta p}_{imeas} \tag{20}$$

$$\overline{\Delta p}_{imeas} = \frac{1}{n}\left(\sum_{i=1}^{n}\Delta p_{i_{meas}}\right) \tag{21}$$

where R is the correlation coefficient, E_{rms} is the root mean square error, F_{rp} is the relative performance factor, $\Delta p_{i_{pred}}$ is the predicted pressure drop, $\Delta p_{i_{meas}}$ is the measured pressure drop, $\overline{\Delta p}_{imeas}$ is average measured pressure drop, $\overline{e}_{imeas}$ is the average change in measured pressure drop, E_6 is the standard deviation, E_5 is the absolute average error, E_4 is the average error, E_3 is the percent standard deviation, E_2 is the absolute average percent error, E_1 is the average percent error, e_i is the actual error, and e_{ri} is the relative error.

The statistical performance analysis of the twenty-eight different Takagi–Sugeno FIS are summarized in Table 5. As shown in Table 5, the FIS with a 0.3 range of influence and trained with normalized dataset achieved the minimum root mean square error of 0.065, minimum relative performance factor of 0.245, and maximum correlation coefficient of 0.902. The optimum FIS consists of 8 inputs, 19 Gaussian input membership functions, 19 rules, 19 linear output membership functions, and one output. Table 5 also clearly shows that for each range of influence, the FIS trained with normalized datasets performed better than the FIS trained with real datasets.

The structure of the developed ANFIS model for FBHP prediction is shown in Fig. 1. The model can be used to predict the FBHP of any production well as long as the eight input parameters used are within the training dataset range (see Table 1). The drawback of the ANFIS model is that it cannot be applied to accurately predict the FBHP of a well data point whose range is not within the training dataset range (see Table 1).

4.5. ANFIS model test

After training and selection of the optimum ANFIS structure, training and testing datasets were applied to the optimum ANFIS structure to evaluate its prediction accuracy. As shown in Figs. 2 and 3, cross-plots with a line of unit slope were used to compare the prediction performance of the proposed ANFIS model based on the training and testing datasets respectively. The closer the cross-plots are to the line of unit slope, the better is the correlation of the predicted FBHP and the measured FBHP [63]. Figs. 2 and 3 clearly show that the developed ANFIS model predicts the FBHP with higher accuracy since most of the cross-plots are very close to the unit slope line. As shown in Fig. 4, the testing dataset, showed a greater deviation from the unit slope line because it was not seen during the training of the model.

4.6. ANFIS white-box model development

The Al-Shammari [8,56] ANFIS model for FBHP prediction was developed and presented without any visible mathematical equation. This made the use of his model on new datasets difficult. Similarly, in section 4.4, this study's ANFIS model for FBHP prediction was developed and presented without any visible mathematical equation (black-box). This also makes its use on new datasets difficult. There is therefore the need to extract the tuned premise and consequence parameter sets of the trained ANFIS black-box model and use them to implement the ANFIS white-box

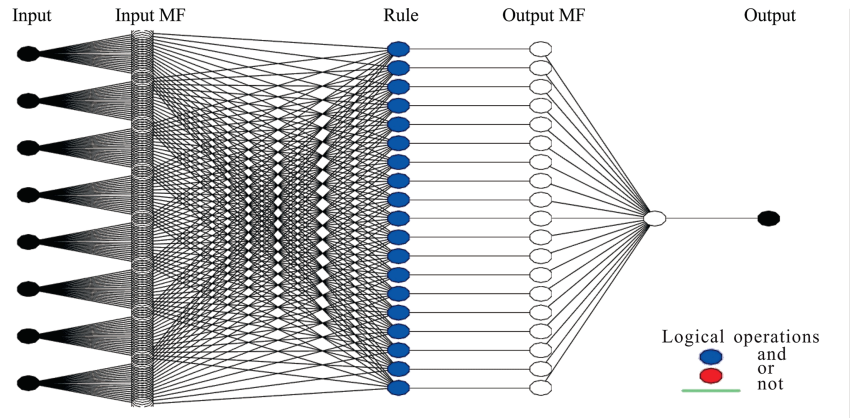

Fig. 1. Developed ANFIS model structure with 8 inputs, 19 IMFs, 19 rules, 19 OMFs, and 1 output.

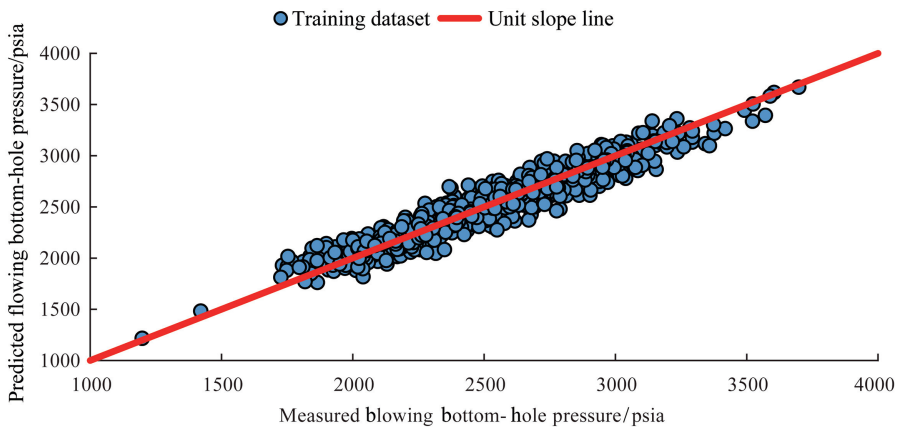

Fig. 2. Measured and predicted FBHP cross-plots with line of unit slope for the training dataset (798 datasets).

model. The extracted consequence and premise parameter sets of the developed ANFIS black-box model are shown in Tables 6 and 7 respectively. The development of the ANFIS white-box model is described next.

First in developing the ANFIS white-box model, the variables $X(1)$, $X(2)$, $X(3)$, $X(4)$, $X(5)$, $X(6)$, $X(7)$, and $X(8)$ were set to represent the input parameters; WHP, WFR, OFR, GFR, WPD, API, ID, and WBHT respectively. Eq. (7) was employed in normalizing the dataset employed in the ANFIS black-box model development to lie between 0 and 1. We followed a similar approach in developing the ANFIS white-box model, where Eq. (7) was used to normalize the inputs to lie between 0 and 1.

In layer 1, the Gaussian membership function (gaussmf) was used to compute the membership degree of each normalized input with the aid of Eq. (22). We followed a similar approach in developing the ANFIS white-box model, where 19 different membership degrees, $\mu(i,j)$ are computed for each normalized input with the aid of the gaussmf equation a defined in Eq. (22).

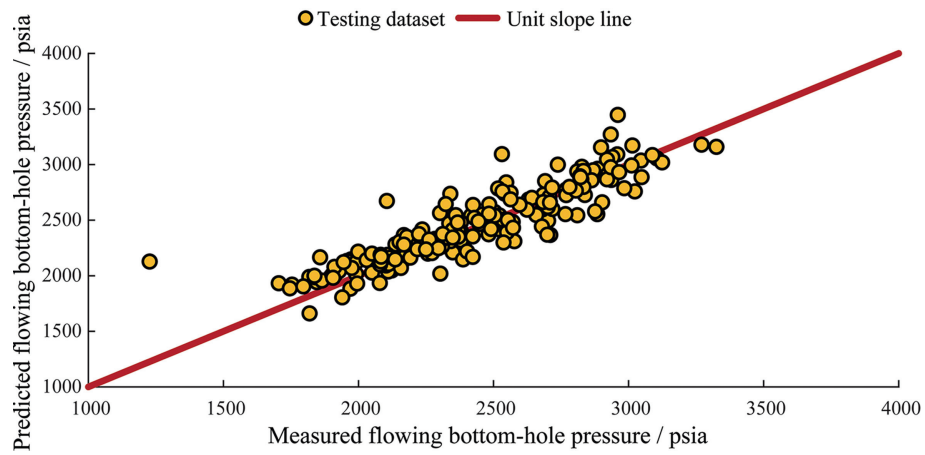

Fig. 3. Measured and predicted FBHP cross-plots with line of unit slope for the testing dataset (203 datasets).

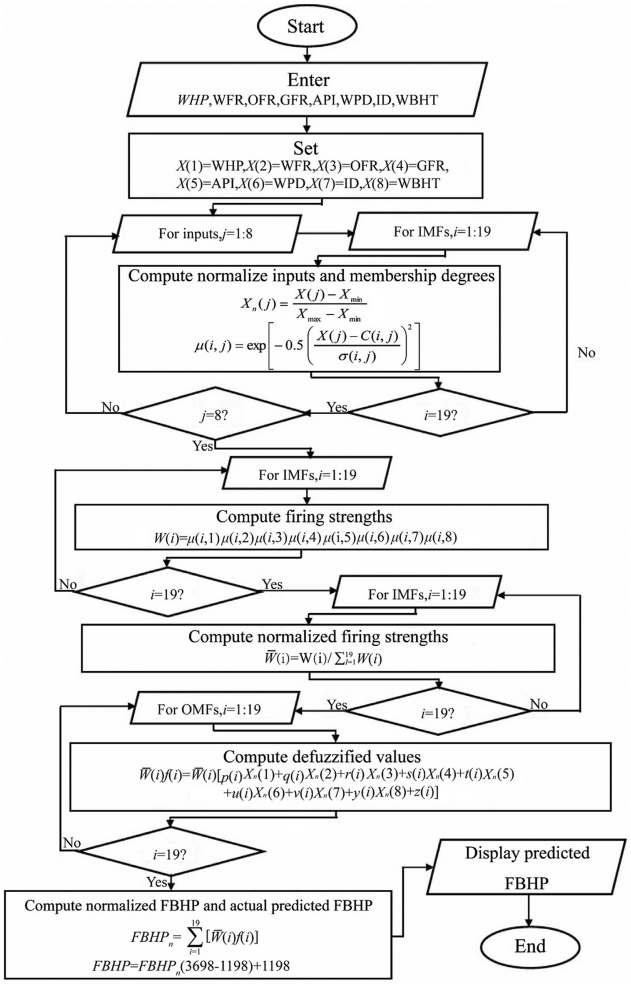

Fig. 4. Flowchart for FBHP prediction with the ANFIS visible mathematical model.

$$\mu(i,j) = \exp\left[-0.5\left(\frac{X_n(j) - C(i,j)}{\sigma(i,j)}\right)^2\right],\ (i = 1:19;\ j = 1:8) \tag{22}$$

where $\mu(i,j)$ is the membership degree for each normalized input $X_n(1)$ through $X_n(8)$, $c(i,1)$ through $c(i,8)$ listed in Table 7 are the ith premise parameter c related to the normalized $X_n(1)$ through $X_n(8)$ respectively, and $\sigma(i,1)$ through $\sigma(i,8)$ listed in Table 7 are the ith premise parameter σ related to the normalized inputs $X_n(1)$ through $X_n(8)$ respectively.

In layer 2, the membership degrees computed in Layer 1 were multiplied to obtain firing strengths for the rules. We followed a similar approach here in developing the ANFIS white-box model, where Eq. (23) was used to compute 19 different firing strengths.

$$W(i) = \mu(i,1)\mu(i,2)\mu(i,3)\mu(i,4)\mu(i,5)\mu(i,6)\mu(i,7)\mu(i,8),\ (i = 1:19) \tag{23}$$

where $W(i)$ is the firing strength of the ith membership function and $\mu(i,1)$ through $\mu(i,8)$ are the ith membership degree for normalized inputs $X_n(1)$ through $X_n(8)$.

In layer 3, the normalized firing strengths were computed by dividing each value of firing strength by the total firing strength. The same approach is followed in the ANFIS white-box model development, where 19 different normalized firing strengths are computed by dividing each of the 19 firing strengths by the sum of the 19 firing strengths computed in layer 2 as given in Eq. (24).

$$\overline{W}(i) = \frac{W(i)}{\sum_{I=1}^{19} W(i)} \tag{24}$$

In layer 4, the normalized firing strengths from layer 3 were defuzzified with the aid of the consequence parameter set. Same approach is followed here as shown in Eq. (25), where normalized firing strengths are defuzzified with the aid of the consequence parameter set as listed in Table 6.

$$\begin{aligned}\overline{W}(i)f(i) =& \overline{W}(i)[p(i)X_n(1) + q(i)X_n(2) + r(i)X_n(3) \\ &+ s(i)X_n(4) + t(i)X_n(5) + u(i)X_n(6) + v(i)X_n(7) \\ &+ y(i)X_n(8) + z(i)],\ (i = 1:19)\end{aligned} \tag{25}$$

where $\overline{W}(i)$ is the normalized firing strength of the ith membership function, $X_n(1), X_n(2), X_n(3), X_n(4), X_n(5), X_n(6), X_n(7)$, and $X_n(8)$ are the normalized WHP_n, WFR_n, OFR_n, GFR_n, WPD_n, API_n, ID_n, and $WBHT_n$ respectively, $p(i)$, $q(i)$, $r(i)$, $s(i)$, $t(i)$, $u(i)$, $v(i)$, and $y(i)$ are the consequence parameters related to the normalized WHP_n, WFR_n, OFR_n, GFR_n, WPD_n, API_n, ID_n, and $WBHT_n$ respectively, and $z(i)$ is the bias of the consequent parameter set.

In layer 5, the defuzzified values are summed up to obtain the normalized FBHP. Same approach is followed here, where the 19 different defuzzified values are summed up as shown in Eq. (26) to obtain the normalized FBHP. Finally, the normalized FBHP was denormalized to lie in the range between the minimum and maximum training FBHPs. A similar approach is followed here in computing the predicted FBHP as shown in Eq. (8).

$$FBHP_n = \sum_{i=1}^{19}[\overline{W}(i)f(i)] \tag{26}$$

The proposed ANFIS visible mathematical FBHP prediction model was also implemented in Microsoft Office Excel and included in the supplementary material. The Excel spreadsheet ANFIS visible mathematic model for FBHP prediction is user-friendly and can easily be used for real-time and accurate predictions of FBHP especially in wellbores that do not have a downhole gauge. The flowchart for predicting FBHP from real-time well data points with the ANFIS visible mathematical model is depicted in Fig. 4.

5. Results and discussion

This study's ANFIS visible mathematical FBHP prediction model is assessed by performing trend, statistical error, and graphical error analyses. First, trend analysis was carried out to investigate if each input parameter's effect on FBHP are simulated accurately by the ANFIS visible mathematical model. Second, statistical and graphical error analyses are used in comparing the predictive performance of the ANFIS visible mathematical model against three mechanistic models, three empirical correlations, one ANN-based model and one ANFIS-based model.

5.1 Trend analysis

As already mentioned, trend analysis was conducted in this study to verify if the ANFIS visible mathematical model correctly simulates the effect of each input parameter on FBHP. This was accomplished by preparing synthetic dataset for each of the input parameters. Each synthetic dataset was prepared by varying one of the input parameters while other input parameters were kept constant. The constant parameters for each synthetic dataset are as given in Table 1, column 6. The trend

Table 6
Extracted consequence parameter set {p, q, r, s, t, u, v, y, and z} of layer 4 used in Eq. (25).

Consequence parameter set									
	WHP_n	WFR_n	OFR_n	GFR_n	WPD_n	API_n	ID_n	$WBHT_n$	BIAS
MF	$p(i)$	$q(i)$	$r(i)$	$s(i)$	$t(i)$	$u(i)$	$v(i)$	$y(i)$	$z(i)$
1	0.838	0.416	0.210	−0.170	0.334	0.895	−1.255	−0.006	0.209
2	0.667	−0.066	0.456	−0.384	0.356	0.052	−0.737	−0.282	0.512
3	0.377	0.205	0.086	−1.044	0.220	0.188	1.093	−0.115	−0.220
4	−0.141	−0.437	2.310	1.397	0.298	0.979	0.136	−0.189	−0.437
5	0.430	0.659	0.594	−0.206	0.465	0.210	−0.324	0.043	−0.146
6	−0.055	−0.086	0.945	−2.488	0.515	−0.327	1.621	0.111	−0.327
7	0.688	0.405	0.249	−0.943	0.518	1.144	−0.124	0.119	−0.299
8	0.401	0.357	−0.044	−1.466	0.133	0.208	0.895	0.221	−0.137
9	0.995	−0.201	1.891	−0.157	0.907	0.321	0.033	−0.416	−0.522
10	0.701	1.952	0.416	−1.830	0.436	0.089	−0.507	0.085	0.226
11	0.735	0.150	−0.379	0.601	0.950	0.054	0.720	0.078	−0.478
12	0.388	0.182	0.330	−0.241	0.431	0.198	−0.158	−0.104	0.187
13	−0.985	0.492	2.096	−2.597	−2.448	−0.979	−1.146	0.842	2.603
14	0.360	0.325	0.045	−0.090	0.511	−0.177	−0.140	−0.025	0.397
15	0.427	0.417	−0.159	−0.137	0.805	−1.046	−1.261	0.524	0.680
16	0.928	0.800	0.001	−0.481	0.721	−0.014	−0.477	0.167	−0.068
17	−0.541	0.761	0.014	−1.050	0.234	0.021	0.067	−0.275	0.884
18	0.397	0.030	−0.038	0.075	1.012	0.788	−0.259	0.789	−0.995
19	0.423	0.477	0.322	−0.309	0.359	−0.335	−0.492	−0.094	0.533

Table 7
Extracted premise parameter set (σ, c) of layer 1 used in Eq. (22).

Premise parameter set																
	WHP		WFR		OFR		GFR		WPD		API		ID		WBHT	
MFs	$\sigma(i,1)$	$c(i,1)$	$\sigma(i,2)$	$c(i,2)$	$\sigma(i,3)$	$c(i,3)$	$\sigma(i,4)$	$c(i,4)$	$\sigma(i,5)$	$c(i,5)$	$\sigma(i,6)$	$c(i,6)$	$\sigma(i,7)$	$c(i,7)$	$\sigma(i,8)$	$c(i,8)$
1	0.107	0.156	0.103	0.132	0.105	0.212	0.105	0.111	0.108	0.460	0.105	0.362	0.099	0.468	0.110	0.789
2	0.107	0.149	0.107	0.160	0.108	0.471	0.104	0.239	0.107	0.505	0.105	0.361	0.098	0.467	0.107	0.789
3	0.106	0.177	0.105	0.012	0.107	0.141	0.106	0.016	0.106	0.230	0.106	0.099	0.106	0.468	0.106	0.039
4	0.105	0.098	0.106	0.436	0.103	0.125	0.103	0.077	0.094	0.499	0.105	0.362	0.103	0.455	0.106	0.792
5	0.107	0.109	0.107	0.038	0.107	0.663	0.106	0.324	0.106	0.472	0.105	0.361	0.112	0.460	0.106	0.789
6	0.106	0.242	0.111	0.218	0.109	0.082	0.105	0.041	0.110	0.478	0.104	0.360	0.100	0.470	0.111	0.787
7	0.106	0.143	0.106	0.001	0.105	0.328	0.106	0.176	0.104	0.494	0.108	0.286	0.116	0.470	0.107	0.566
8	0.110	0.164	0.107	0.000	0.104	0.090	0.106	0.014	0.108	0.439	0.107	0.100	0.106	0.468	0.106	0.040
9	0.110	0.096	0.105	0.358	0.106	0.366	0.108	0.207	0.106	0.521	0.108	0.363	0.107	0.459	0.107	0.790
10	0.107	0.421	0.107	0.082	0.106	0.125	0.106	0.062	0.111	0.521	0.101	0.358	0.103	0.463	0.097	0.791
11	0.107	0.099	0.110	0.794	0.107	0.213	0.107	0.148	0.106	0.556	0.106	0.362	0.106	0.459	0.106	0.789
12	0.106	0.184	0.107	0.178	0.107	0.686	0.106	0.400	0.106	0.567	0.105	0.539	0.105	0.468	0.108	0.894
13	0.107	0.153	0.104	0.452	0.107	0.019	0.108	0.022	0.113	0.572	0.108	0.363	0.113	0.471	0.111	0.787
14	0.100	0.545	0.107	0.051	0.103	0.413	0.106	0.168	0.100	0.498	0.104	0.539	0.108	0.459	0.111	0.894
15	0.104	0.142	0.105	0.435	0.106	0.314	0.106	0.201	0.105	0.536	0.107	0.538	0.101	0.468	0.106	0.895
16	0.108	0.103	0.107	0.167	0.108	0.056	0.107	0.023	0.106	0.489	0.108	0.363	0.106	0.104	0.106	0.790
17	0.106	0.443	0.111	0.309	0.105	0.078	0.105	0.068	0.102	0.538	0.111	0.366	0.118	0.463	0.088	0.796
18	0.105	0.095	0.108	0.681	0.106	0.418	0.107	0.308	0.106	0.554	0.106	0.538	0.106	0.459	0.106	0.895
19	0.109	0.423	0.103	0.032	0.107	0.368	0.107	0.180	0.108	0.394	0.109	0.363	0.098	0.467	0.108	0.789

analysis was carried out to investigate the effects of each of the eight input parameters on FBHP. For qualitative comparison, the trend analysis was performed for the proposed ANFIS visible mathematical model and one industry standard method for multiphase pipe flow modelling, e.g., modified Beggs & Brill [29] correlation.

WHP effect on FBHP is as shown in Fig. 5, where FBHP is increased following an increase in WHP for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This trend is justified by the general equation for the computation of FBHP from WHP as given in Eq. (26a). As shown in Eq. (26a), the FBHP is directly proportional to the WHP, which implies that an increase in WHP will increase FBHP. This proves that the proposed ANFIS visible mathematical model correctly simulates the increase in FBHP with an increase in WHP.

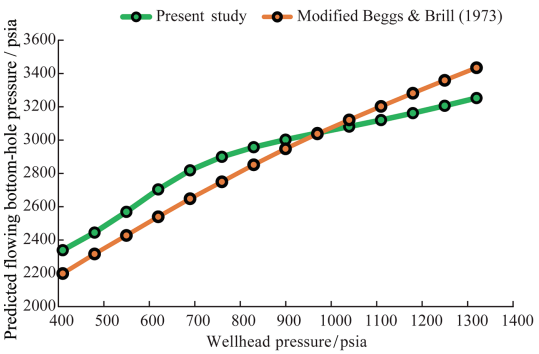

Fig. 5. Trend of FBHP for varying values of WHP.

$$p_{wf} = p_{wh} + \sum_{i=1}^{n}\left[\left(\frac{dp}{dl}\right)_i L_i\right] = p_{wh} + \Delta p_T \tag{26a}$$

where p_{wf} is the FBHP, p_{wh} is the WHP, $\left(\frac{dp}{dL}\right)_i$ is the pressure gradient for each tubing segment, Δp_T is the total pressure drop (TPD), and L_i is the ith tubing segment length.

WFR effect on FBHP is as shown in Fig. 6, where FBHP is increased as expected following an increase in WFR for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This trend is justified by the equations for the computation of no-slip liquid holdup (input liquid content), no-slip density, total pressure gradient, and FBHP as given by Eqs. (26)–(29) respectively. These equations clearly show that increasing WFR results in an indirect increase in FBHP. Fig. 6 and Eqs. (26)–(29) clearly show that the proposed ANFIS visible mathematical model correctly simulates the increase in FBHP with an increase in WFR.

$$\lambda_L = \frac{q_L}{q_L + q_g} = \frac{q_o + q_w}{q_g + q_o + q_w} \tag{27}$$

$$\rho_{mn} = \lambda_L \rho_L + (1 - \lambda_L)\rho_g \tag{28}$$

$$\left(\frac{dp}{dL}\right)_i = \frac{g}{g_c}\rho_{mn}\sin\theta + \frac{f_{2F}\rho_{mn}v_m^2}{2g_c d} \tag{29}$$

where λ_L is the no-slip liquid holdup, q_o is the oil flow rate, q_w is the water flow rate, q_g is the gas flow rate, q_L is the liquid flow rate, ρ_g is the gas density, ρ_L is the liquid density, ρ_{mn} is the no-slip mixture density, v_m is the mixture velocity, f_{2F} is the two-phase friction factor, and d is the tubing diameter.

The effect of OFR on FBHP is as shown in Fig. 7, where FBHP is increased following an increase in OFR for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This trend is also justified by the no-slip liquid holdup, no-slip density, ith tubing segment pressure gradient, and FBHP equations as given by Eqs. (26)–(29) respectively. It is evident from these equations that FBHP is increased indirectly following an increase in OFR. Hence, Fig. 7 and Eqs. (26)–(29) clearly show that the proposed ANFIS visible mathematical model correctly simulates the anticipated increase in FBHP following an increase in OFR.

GFR effect on FBHP is as shown in Fig. 8, where the FBHP is decreased following an increase in GFR for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This trend is justified by Eqs. (26)–(29) used for computation of no-slip liquid holdup, no-slip density, ith tubing segment pressure gradient, and FBHP respectively. It is evident from these equations that FBHP is inversely proportional to GFR. Fig. 8 and Eqs. (26)–(29) indicate that the proposed ANFIS visible mathematical model correctly simulates the anticipated decrease in FBHP as GFR increases.

WPD effect on FBHP is as shown in Fig. 9, where FBHP is increased as WPD increases for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This trend is justified by the equation for the computation of FBHP as given by Eq. (26). Remember that WPD is the algebraic sum of L_i and that L_i is increased following an increase in WPD. As shown in Eq. (26), L_i is directly proportional to FBHP. Hence, Fig. 9 and Eq. (26) show that the proposed ANFIS visible mathematical model correctly simulates the expected increase in FBHP following an increase in WPD.

API effect on FBHP is as shown in Fig. 10, where FBHP increases slightly to a maximum value and then decreases sharply as API

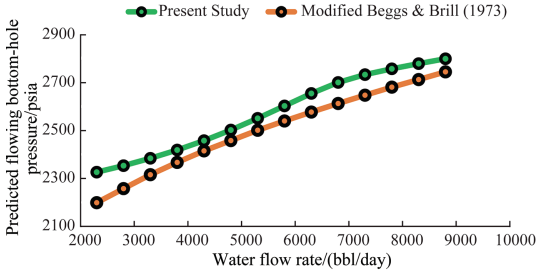

Fig. 6. Trend of FBHP for varying values of WFR.

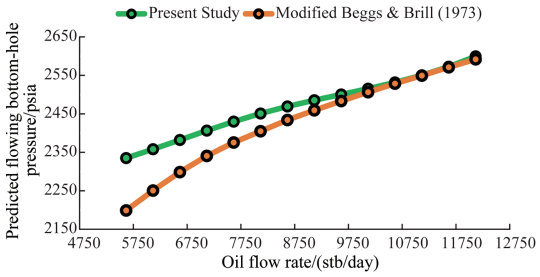

Fig. 7. Trend of FBHP for varying values of OFR.

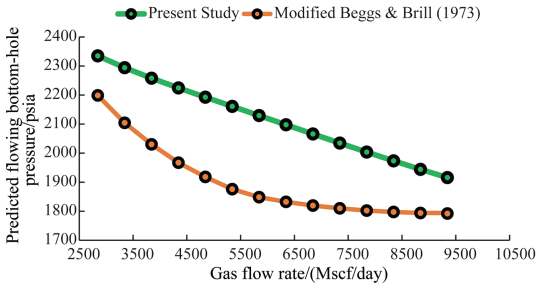

Fig. 8. Trend of FBHP for varying values of GFR.

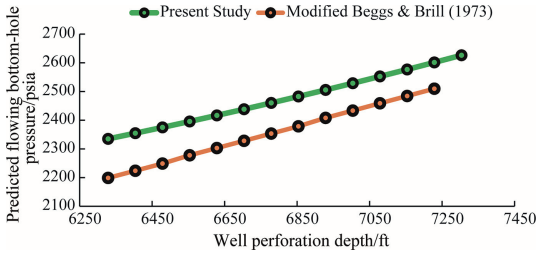

Fig. 9. Trend of FBHP for varying values of WPD.

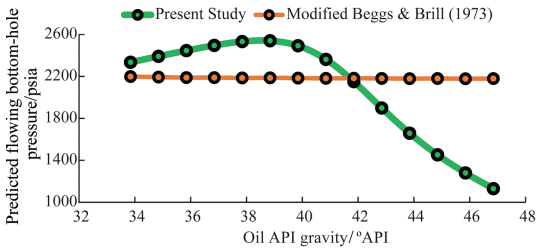

Fig. 10. Trend of FBHP for varying values of API.

increases further. The sharp decrease in FBHP from a maximum API of 38.85 is justified by the fact that at an API greater than 39, a high percent of lighter hydrocarbons are dissolved in the oil. This will aerate the fluid column with a resulting sharp decrease in FBHP as API increases. This clearly shows that the proposed ANFIS visible mathematical model correctly simulates the expected initial slight increase in FBHP up to a maximum value followed by a sharp decrease in FBHP with a further increase in API. As shown in Fig. 10, FBHP remained approximately constant with increase in API for the industry standard modified Beggs & Brill [29] correlation.

ID effect on FBHP is as shown in Fig. 11, where FBHP is decreasing with increase in ID for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This trend is justified by Eqs. (26)–(29) used for the computation of the ith tubing segment pressure gradient and FBHP respectively. It is evident from Eq. (29) that the ith tubing segment pressure gradient is inversely proportional to ID. It is also evident from Eq. (26) that FBHP is directly proportional to the ith tubing segment pressure gradient. These imply that FBHP is inversely proportional to ID. Hence, Fig. 11 and Eqs. (26)–(29) clearly show that the proposed ANFIS visible mathematical model correctly simulates the expected decrease in FBHP as ID increases.

WBHT effect on FBHP is as shown in Fig. 12, where FBHP is decreasing with increase in WBHT for the proposed ANFIS visible mathematical model and the industry standard multiphase flow modelling correlation. This is justified first by the fact that WBHT is inversely proportional to liquid holdup. Second, Eq. (28) clearly shows that mixture density is directly proportional to liquid holdup. Third, Eq. (29) also shows, the ith tubing segment pressure gradient is directly proportional to mixture density. Lastly, Eq. (26) show that FBHP is directly proportional to TPD. These clearly show that WBHT is inversely proportional to FBHP. It is clear from the above and Fig. 12 that the proposed ANFIS visible mathematical model correctly simulates the expected decrease in FBHP as WBHT increases.

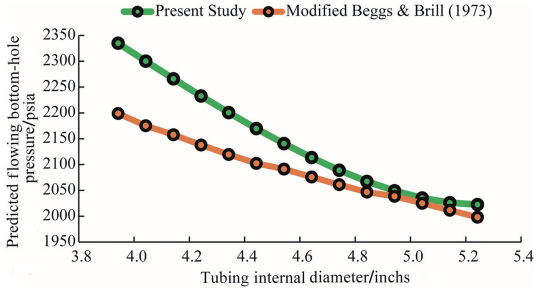

Fig. 11. Trend of FBHP for varying values of ID.

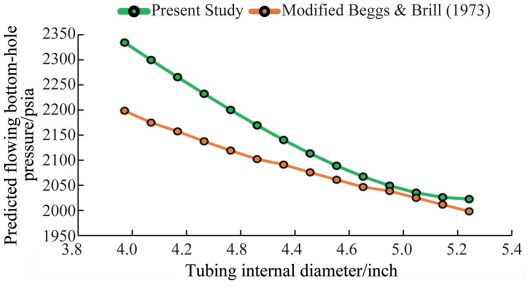

Fig. 12. Trend of FBHP for varying values of WBHT.

5.2. Comparison of ANFIS model against existing models

As already mentioned, 206 test data points were used to compare the prediction performance of the ANFIS visible mathematical model against three empirical correlations [13,28,29], three mechanistic models [2,36,39], one ANN-based model [62], and one ANFIS-based model [8,56]. The comparison was performed with the aid of statistical and graphical error analyses.

5.3. Statistical error analysis

Statistical error analysis was employed in performing the comparative evaluation of the proposed ANFIS model against existing mechanistic models, empirical correlations, and machine learning models (ANN and ANFIS) for FBHP prediction in wellbores. The statistical error analysis is based on six statistical error parameters and the relative performance factor [2,86]. It is important that we briefly define these six statistical error parameters and the relative performance factor employed in this study.

The average percent error, E_1, is given mathematically by Eq. (14) and expresses the overall performance trend with respect to the observed drop in pressure [2]. A positive E_1 indicates that the pressure drop was over-predicted, whereas a negative E_1 indicates that it was under-predicted. Absolute average percent error, E_2, is given mathematically by Eq. (15) and expresses the magnetude of the errors on average in relation to the observed drop in pressure. E_2 is a more important statistical error parameter than E_1 since it prevents the positive relative error and the negative relative error from cancelling each other out [3,4,87]. Percent standard deviation, E_3, also given mathematically by Eq. (16), represents the magnitude of the error scattering relative to the observed drop in pressure [2]. Average error, E_4, defined by Eq. (17), expresses the overall performance trend without regard for the observed drop in pressure [3,4]. A positive E_4 indicates that the pressure drop was over-predicted, whereas a negative E_4 indicates that it was under-predicted [87]. Absolute average error, E_5, defined by Eq. (18), expresses the magnetude of the errors on average irrespective of the observed drop in pressure and prevents the positive actual error and the negative actual error from cancelling each other out [2,87]. Standard deviation, E_6, given in Eq. (19), expresses the degree of error scattering irrespective of the observed drop in pressure [2].

Relative performance factor, F_{rp}, given mathematically by Eq. (9) takes into account the impact of all six statistical error parameters (E_1, E_2, E_3, E_4, E_5, and E_6). F_{rp} was recommended by Ansari et al, [2] for conducting comparison among a group of models with minimum and maximum values of zero and six for the best-performing model and worst-performing model respectively.

Table 8 summarizes the results of the statistical error analysis employed in the current study in performing the comparative evaluation of this study's ANFIS visible mathematical model, three mechanistic models, three empirical correlations, one ANN-based model and one ANFIS-based model for predicting FBHP in wellbores. Column 1 reports the different models considered for the cooperative study. The statistical error parameters E_1 through E_6 are reported in columns 2 through 7 while the relative performance factor, F_{rp} is reported in column 8. The lowest values of E_1 through E_6 and F_{rp} are represented by blue cells. As shown in Table 8, the different models are organized in ascending sequence of F_{rp} values. As shown in Table 8, the lowest values of E_2 (3.71%), E_4 (1.89 psia) and E_5 (75.19 psia) were achieved by this study's ANFIS visible mathematical model. The lowest values of E_1 (0.04%) and E_3 (0.98%) were achieved by the Al-Shammari [8,56] ANFIS model and lowest value of E_6 (851.63 psia) by the ANN model developed by Tariq et al. [62]. Similarly, largest values of E_1, E_2, E_3, E_4, and E_5 were achieved

Table 8
Statistical performance evaluation of this study's ANFIS visible mathematical FBHP prediction model and existing FBHP prediction models based on the Ayoub [63] 206 well data points.

Model	E_1 (%)	E_2 (%)	E_3 (%)	E_4 (psia)	E_5 (psia)	E_6 (psia)	F_{rp} (-)
This study	0.51	**3.71**	7.34	**1.89**	**75.19**	1080.57	**0.09**
Tariq et al. [62]	2.65	3.72	37.81	57.49	80.45	**851.63**	0.35
Al-Shammari [8,56]	**0.04**	5.60	**0.98**	−14.70	115.79	1650.24	0.39
Modified Hagedorn & Brown [28]	−1.46	8.08	20.80	−21.43	171.01	2444.14	0.96
Modified Beggs & Brill [29]	−7.90	8.87	112.55	−169.48	188.28	1760.84	1.75
Poettmann & Carpenter [13]	−5.18	14.17	73.77	−119.99	310.34	4192.94	2.50
Petalas & Aziz [39]	−13.92	15.01	198.31	−305.32	325.04	3169.59	3.49
Hasan & Kabir [36]	−16.19	19.43	230.56	−362.17	423.28	5359.15	4.74
Ansari et al. [2]	−22.76	23.38	324.23	−501.33	511.99	5116.56	5.95

by the mechanistic model developed by Ansari et al. [2] and the largest value of E_6 was achieved by mechanistic model developed by Hasan & Kabir [36].

In order to demonstrate the robustness of the developed ANFIS model, the relative performance factor, F_{rp} was adopted because it incorporates the effects of E_1, E_2, E_3, E_4, E_5 and E_6 as defined in Eq. (9). Table 8 clearly shows that this study's ANFIS model with an F_{rp} of 0.09 performed best while the Ansari et al. [2] model with an F_{rp} of 5.95 performed worst. As shown in Table 8, the machine learning models performed best followed by the empirical correlations with the mechanistic models performing worst. A separate study conducted by Ref. [44] revealed that the mechanistic models performed worst because of the severe underprediction of the slug flow and churn flow mechanistic models.

5.4. Graphical error analysis

Cross-plots and residual plots are the two graphical error analysis methods used in this study for the comparative performance evaluation of the novel ANFIS visible mathematical model against the current empirical correlations, mechanistic models, ANN-based model, and ANFIS-based model.

5.5. Cross-plots

Cross-plots of the pressure drop predicted by the ANFIS visible mathematical model, empirical correlations, mechanistic models, ANN-based model, and ANFIS-based model on the vertical axis against the observed drop in pressure on the horizontal axis are shown in Fig. 13. A unit slope line, which represents a line of perfect correlation, is also drawn. Furthermore, deviation lines of +20% and −20%, which respectively show over-prediction and under-prediction by the models, were also drawn. The best-performing model should (1) have majority of the cross-plots on or very close to the line of unit slope and (2) have majority of the cross-plots within the deviation lines of +20% and −20%. The reverse is true for the worst-performing model.

Fig. 13 clearly shows that majority of the cross-plots are on or very close to the line of unit slope for the current ANFIS visible mathematical model and hence considered the best-performing model. Also, 99% of the cross-plots are within the deviation lines of +20% and −20% for the current ANFIS visible mathematical model and Tariq et al. [62] model. Fig. 13 also shows that only 50% of the cross-plots are within the deviation lines of +20% and −20% for the model developed by Ansari et al. [2] and hence considered the worst-performing model. The Al-Shammari [8,56] model, modified Hagedorn & Brown [28] correlation, modified Beggs & Brill [29] correlation, Poettmann & Carpenter [13], correlation Petalas & Aziz [39], model, and Hasan & Kabir [36] model had 98%, 92%, 90%, 75%, 70% and 60% of cross-plots within the deviation lines of +20% and −20%.

5.6. Residual plots

Residual plots for the developed ANFIS visible mathematical model, empirical correlations, mechanistic models, ANN-based model, and ANFIS-based model are shown in Fig. 14. The best-performing correlation or model should have the most number of residual plots within the +300 psia and −300 psia lines of residual deviation. The reverse is true for the worst-performing correlation or model.

It is clear from Fig. 14 that this study's ANFIS visible mathematical model and the Tariq et al. [62] ANN visible mathematical model showed the best performance since 98% of their respective residual plots are within the +300 psia and −300 psia lines of residual deviation. The Ansari et al. [2] mechanistic model showed

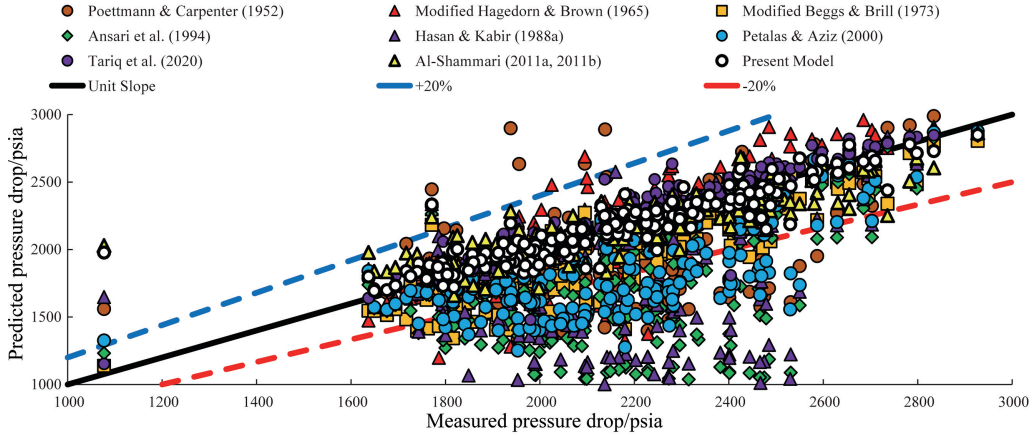

Fig. 13. Cross-plots of pressure drop for the developed ANFIS visible mathematical model, empirical correlations, mechanistic models, ANN-based model, and ANFIS-based model.

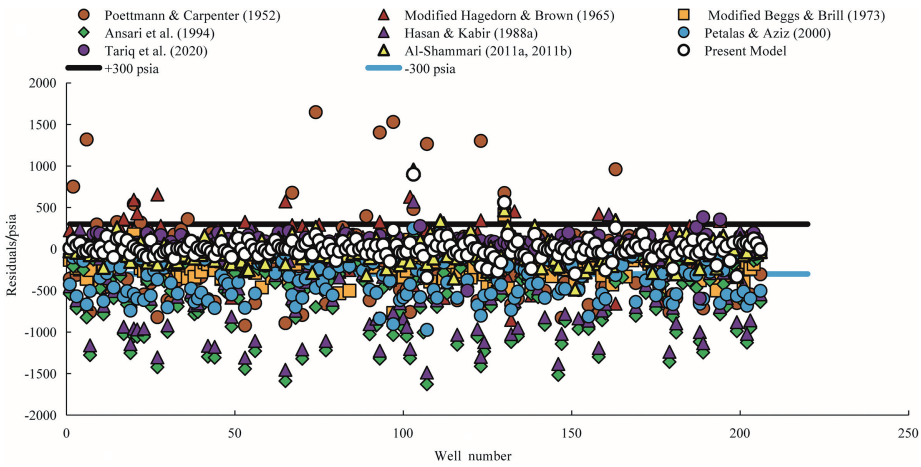

Fig. 14. Residual plots for the developed ANFIS visible mathematical model, empirical correlations, mechanistic models, ANN-based model, and ANFIS-based model.

the worst performance with 37% of residual plot within the +300 psia and −300 psia lines of residual deviation. In addition, Fig. 14 reveals that 51%, 52%, 59%, 80%, 85%, and 96% residual plots for the Hasan & Kabir [36] model Petalas & Aziz [39], model Poettmann & Carpenter [13] correlation, modified Beggs & Brill [29] correlation, modified Hagedorn & Brown [28] correlation, and Al-Shammari [8,56] ANFIS model respectively are within the +300 psia and −300 psia lines of residual deviation.

6. Conclusions and recommendations

6.1. Conclusions

Adaptive neuro-fuzzy inference system (ANFIS) coupled with hybrid learning (HL) optimization method was successfully used in this study to develop a visible mathematical model for prediction of FBHP in wellbores in real-time.

Based on the findings and analysis included in this study, the following conclusions can be made.

(1) Statistical error analysis performed on 28 different Takagi–Sugeno fuzzy inference systems (FIS) showed that a FIS with a 0.3 range of influence and trained with a normalized dataset achieved the best FBHP prediction performance.
(2) The optimal ANFIS black-box model was then translated into the ANFIS white-box model with the Gaussian input and the linear output membership functions and the extracted tuned premise and consequence parameter sets.
(3) Trend analysis conducted revealed that the anticipated effect of various inputs on FBHP were correctly simulated by the proposed ANFIS visible mathematical model. This is evident in (i) an increase in FBHP with an increase in WHP, WFR, OFR, and WPD respectively and (ii) a drop in FBHP following an increase in API, ID, WBHT, and GFR respectively.
(4) The minimum values of relative performance factor (F_{rp} = 0.09), absolute average percent error (E_2=3.71%), average error (E_4=1.89 *psia*), and absolute average error (E_5= 75.19 *psia*) were achieved by novel ANFIS visible mathematical model. Also, the minimum value of standard deviation (E_6=851.63 *psia*) was achieved by Tariq et al. [62] ANN visible mathematical model. In addition, the Al-Shammari [8,56] ANFIS model achieved the minimum values of the average percent error (E_1=0.04%) and the percent standard deviation (E_3=0.98%).
(5) Majority of the pressure drop cross-plots for the current ANFIS visible mathematical model are on or very close to the line of unit slope with 99% of the cross-plots within the +20% and −20% deviation lines.
(6) This study's ANFIS visible mathematical model and the Tariq et al. [62] ANN visible mathematical model showed best performance with 98% of residual plots within the +300 psia and −300 psia lines of residual deviation.
(7) This study's ANFIS visible mathematical model has a wider range of applicability than the ANFIS-based and ANN-based models developed by Al-Shammari [8,56] and Tariq et al. [62] respectively.
(8) The advantage of this study's ANFIS visible mathematical model is that it can be employed in new wells for FBHP prediction without the need for expensive machine learning frameworks such MATLAB fuzzy logic toolbox for training purposes.
(9) The drawback of this study's ANFIS visible mathematical model is that it cannot be applied to accurately predict the FBHP of a data point whose range is not within the training dataset range as shown in Table 1.
(10) There is a risk of local minimum with the derivative based hybrid optimization algorithm used in the current study.

6.2. Recommendations

(1) The novel ANFIS model should only be used when the input parameter ranges are within the training dataset input parameter ranges as defined in Table 1.
(2) New training datasets covering wider input parameter ranges should be added to the original training datasets and the resulting dataset used to train the ANFIS. This will improve the model's range of applicability and accuracy.
(3) Training the ANFIS with heuristic-derivative based hybrid optimization algorithms is recommended and the performance result compared with this study's performance.

Nomenclature

R	Correlation coefficient
E_{rms}	Root mean square error
$\Delta p_{i_{pred}}$	Predicted pressure drop, psia

$\Delta p_{i_{meas}}$	Measured pressure drop, psia
$\overline{\Delta p_{imeas}}$	Average measured pressure drop, psia
$\overline{e}_{imeas}$	Average change in measured pressure drop, psia
F_{rp}	Relative performance factor
E_6	Standard deviation, psia
E_5	Absolute average error, psia
E_4	Average error, psia
E_3	Percent standard deviation, %
E_2	Absolute average percent error, %
E_1	Average percent error, %
e_i	ith data point's actual error
e_{ri}	ith data point's relative error
$\{A_i, B_i\}$	Linguistic label to node i
w_i	Firing strength
$\overline{w_i}$	Normalized firing strength from layer 3
df	Degree of freedom
n	Number of data points
$X_{n(0:1)}$	Normalized value between 0 and 1 of the parameter
X	Original parameter value
X_{min}	Minimum parameter value as shown in column 4 of Table 1
X_{max}	Maximum parameter value as column 5 of Table 1
$\mu(i,j)$	Membership degree for each normalized input
$W(i)$	Firing strength of the ith membership function
$\overline{W}(i)$	Normalized firing strength of the ith membership function
$\{p, q, r, s, t, u, v, y,$ and $z\}$	Consequence parameter set.
(σ, c)	Premise parameter set
p_{wf}	Flowing bottom-hole pressure, psia
p_{wh}	Wellhead Pressure, psia
$\left(\frac{dp}{dL}\right)_T$	Total pressure gradient (TPG), psi/foot
L	Well perforation depth, feet
d	Tubing diameter, inch
v_m	Mixture velocity, ft/sec
f_{2F}	Two-phase friction factor
ρ_g	Gas density, lb/cuft
ρ_L	Liquid density, lb/cuft
ρ_{mn}	No-slip mixture density, lb/cuft
q_w	Water volumetric flow rate, ft^3/sec
q_o	Oil volumetric flow rate, ft^3/sec
q_L	Liquid volumetric flow rate, ft^3/sec
q_g	Gas volumetric flow rate, ft^3/sec
λ_L	No-slip liquid holdup
$FBHP_n$	Normalized flowing bottom-hole pressure
$WBHT_n$	Normalized well bottom-hole temperature
ID_n	Normalized internal diameter of the tubing
API_n	Normalized oil API gravity
WPD_n	Normalized well perforation depth
GFR_n	Normalized gas flow rate
OFR_n	Normalized oil flow rate
WFR_n	Normalized water flow rate
WHP_n	Normalized wellhead pressure
FBHP	Flowing bottom-hole pressure, psia
WHT	Wellhead temperature, °F
WBHT	Well bottom-hole temperature, °F
ID	Internal diameter of the tubing, inch
API	Oil API gravity, °API
WPD	Well perforation depth, feet
GFR	Gas flow rate, Mscf/d
OFR	Oil flow rate, stb/d
WFR	Water flow rate, stb/d
WHP	Wellhead pressure, psia
ANN	Artificial neural network
ANFIS	Adaptive neuro-fuzzy inference system
ML	Machine learning
FIS	Fuzzy inference system
PSO	Particle swarm optimization
HGA	Hybrid genetic algorithm
FLC	Fuzzy logic controller
WOA	Whale optimization algorithm
KHA	Krill herd algorithm
FA	Firefly algorithm
BCO	Bee colony optimization
DFA	Dragonfly algorithm
SFLA	Shuffled frog-leaping algorithm
IWO	Invasive weed optimization
GWO	Grey wolf optimization
DE	Differential evolution
GA	Genetic algorithm
ABC	Artificial bee colony
APSO	Accelerated particle swarm optimization
RCGA	Real-coded genetic algorithm
aABC	Hybrid artificial bee colony algorithm
HL	Hybrid learning
BP/GD	backpropagation gradient descent
LSE	Least square estimation
RLSE	Recursive least square estimation
LM	Levenberg-Marquardt
KF	Kalman filter
EKF	Extended Kalman filter
TCC	Threshold correlation coefficient
RF	Random forest
KNN	K-Nearest neighbours
CC	Correlation coefficient
IMF	Input membership functions
OMF	Output membership functions
MF	Membership functions
gaussmf	Gaussian membership function

References

[1] E.A. Kanin, A.A. Osiptsov, A.L. Vainshtein, E.V. Burnaev, A predictive model for steady-state multiphase pipe flow: machine learning on lab data, J. Petrol. Sci. Eng. 180 (May) (2019) 727–746, https://doi.org/10.1016/j.petrol.2019.05.055.

[2] A.M. Ansari, N.D. Sylvester, U. Akron, C. Sarica, O. Shoham, A comprehensive mechanistic model for upward two-phase flow in wellbores, SPE Prod. Facil. 9 (2) (1994) 143–152, https://doi.org/10.2118/20630-PA.

[3] A.S. Kaya, C. Sarica, J.P. Brill, Comprehensive Mechanistic Modeling of Two-phase Flow in Deviated Wells, SPE Annual Technical Conference and Exhibition, Houston, Texas, 1999, pp. 3–6, https://doi.org/10.2118/56522-MS.

[4] A.S. Kaya, C. Sarica, J.P. Brill, Mechanistic modeling of two-phase flow in deviated wells, SPE Prod. Facil. 16 (3) (2001) 156–165, https://doi.org/10.2118/72998-PA.

[5] E. Eltahan, R. Ganjdanesh, W. Yu, K. Sepehrnoori, R. Williams, J. Nohavitsa, Machine learning approach to improve calculated bottom-hole pressure, in: SPE/AAPG/SEG Unconventional Resources Technology Conference, 1–21, 2021, https://doi.org/10.15530/urtec-2021-5645.

[6] A.E. Dukler, M.G. Hubbard, A model for gas–liquid slug flow in horizontal and near horizontal tubes, Ind. Eng. Chem. Fundam. 14 (4) (1975) 337–347, https://doi.org/10.1021/i160056a011.

[7] A. Olufemi, F.A.S. Adesina, F. Olugbenga, Predictive tool for bottom-hole pressure in multiphase flowing wells, Petrol. Coal 50 (3) (2008) 67–73.

[8] A. Al-Shammari, Accurate prediction of pressure drop in two-phase vertical flow systems using artificial intelligence, in: SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition, 2011, https://doi.org/10.2118/149035-ms.

[9] M. Awadalla, H. Yousef, A. Al-Shidani, A. Al-Hinai, Artificial intelligent techniques for flow bottom hole pressure prediction, Int. J. Comput. Inf. Technol. Eng. 15 (12) (2016) 7263–7283.

[10] J.V. Vogel, Inflow performance relationships for solution-gas drive wells, J. Petrol. Technol. 243 (1968) 83–92.

[11] U.I. Duru, O.I. Nwanwe, C.C. Nwanwe, A.O. Arinkola, A.O. Chikwe, Evaluating lift systems for oil wells using integrated production modeling : a case study of A Niger delta, J. Petrol. Eng. Technol. 11 (3) (2021) 32–48, https://doi.org/10.37591/jopet.v11i3.6057.

[12] C.C. Nwanwe, U.I. Duru, O.I. Nwanwe, A.O. Chikwe, K.T. Ojiabo, C.T. Umeojiakor, Optimum tubing size prediction model for vertical multiphase flow during flow production period of oil wells, J. Pet. Explor. Prod. Technol. 10 (7) (2020) 2989–3005, https://doi.org/10.1007/s13202-020-00964-8.
[13] F.H. Poettmann, P.G. Carpenter, The multiphase flow of gas , oil , and water through vertical flow strings with application to the design of gas-lift installations, API Drill. Prod. Pract. (1952) 257–317.
[14] B.C. Craft, M.F. Hawkins, Applied Petroleum Reservoir Engineering, second ed., Prentice Hall PTR, 1991.
[15] J.S. Kim, L.W. Lake, T.F. Edgar, Integrated capacitance-resistance model for characterizing waterflooded reservoirs, IFAC Proc. Vol. 1 (PART 1) (2012) 19–24, https://doi.org/10.3182/20120531-2-NO-4020.00009.
[16] S.H. Yousefi, F. Rashidi, M. Sharifi, M. Soroush, Prediction of immiscible gas flooding performance: a modified capacitance–resistance model and sensitivity analysis, Petrol. Sci. 16 (5) (2019) 1086–1104, https://doi.org/10.1007/s12182-019-0342-6.
[17] A.A. Yousef, P. Gentil, J.L. Jensen, L.W. Lake, A capacitance model to infer interwell connectivity from production- and injection-rate fluctuations, SPE Reservoir Eval. Eng. 9 (6) (2006) 630–646, https://doi.org/10.2118/95322-pa.
[18] M. Sayarpour, E. Zuluaga, C.S. Kabir, L.W. Lake, The use of capacitance-resistance models for rapid estimation of waterflood performance and optimization, J. Petrol. Sci. Eng. 69 (3–4) (2009) 227–238, https://doi.org/10.1016/j.petrol.2009.09.006.
[19] D. Weber, T.F. Edgar, L.W. Lake, L. Lasdon, S. Kawas, M. Sayarpour, Improvements in capacitance-resistive modeling and optimization of large scale reservoirs, in: SPE Western Regional Meeting 2009 - Proceedings, 2009, pp. 369–385, https://doi.org/10.2118/121299-ms.
[20] A.P. Nguyen, J.S. Kim, L.W. Lake, T.F. Edgar, B. Haynes, Integrated capacitance resistive model for reservoir characterization in primary and secondary recovery, in: Proceedings - SPE Annual Technical Conference and Exhibition, 5(2006), 2011, pp. 4162–4181, https://doi.org/10.2118/147344-ms.
[21] D. Wang, Y. Li, J. Zhang, C. Wei, Y. Jiao, Q. Wang, Improved CRM model for inter-well connectivity estimation and production optimization: case study for karst reservoirs, Energies 12 (5) (2019), https://doi.org/10.3390/en12050816.
[22] R. W. d Holanda, E. Gildin, J.L. Jensen, L.W. Lake, C.S. Kabir, A state-of-the-art literature review on capacitance resistance models for reservoir characterization and performance forecasting, Energies 11 (12) (2018), https://doi.org/10.3390/en11123368.
[23] X. Wang, J. Wang, Z. Xia, Improved capacitance model considering bottom-hole flowing pressure and interference between oil wells, Energy Explor. Exploit. 38 (6) (2020) 2277–2295, https://doi.org/10.1177/0144598720918401.
[24] J.P. Brill, H. Mukherjee, Multiphase Flow in Wells, Society of Petroleum Engineers, Richardson, TX, 1999.
[25] P.B. Baxendell, R. Thomas, The calculation of pressure gradients in high-rate flowing wells, J. Petrol. Technol. 13 (10) (1961) 1023–1028, https://doi.org/10.2118/2-pa.
[26] G.H. Fancher, K.E. Brown, Prediction of pressure gradients for multiphase flow in tubing, Soc. Petrol. Eng. J. 3 (1) (1963) 59–69, https://doi.org/10.2118/440-pa.
[27] H. Asheim, Mona, an accurate two-phase well flow model based on phase slippage, SPE Prod. Eng. 1 (3) (1986) 221–230, https://doi.org/10.2118/12989-PA.
[28] A.R. Hagedorn, K.E. Brown, Experimental study of pressure gradients occurring during continuous two-phase flow in small-diameter vertical conduits, J. Petrol. Technol. 17 (4) (1965) 475–484, https://doi.org/10.2118/940-pa.
[29] H.D. Beggs, J.P. Brill, A Study of two-phase flow in inclined pipes, J. Petrol. Technol. 25 (5) (1973) 607–617, https://doi.org/10.2118/4007-PA.
[30] H. Duns Jr., N.C.J. Ros, Vertical flow of gas and liquid mixtures in wells, in: Proceedings of the 6th World Petroleum Congress, 19–26 June, Frankfurt Am Main, Germany, 1963, pp. 451–465.
[31] J. Orkiszewski, Predicting two-phase pressure drops in vertical pipe, J. Petrol. Technol. 19 (6) (1967) 829–838, https://doi.org/10.2118/1546-pa.
[32] A.U. Yahaya, A. Al Gahtani, A comparative study between empirical correlations & mechanistic models of vertical multiphase flow, in: SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition, 2010, https://doi.org/10.2118/136931-ms.
[33] L.E. Gomez, O. Shoham, Z. Schmidt, R.N. Chokshi, T. Northug, Unified mechanistic model for steady-state two-phase flow: horizontal to vertical upward flow, SPE J. 5 (3) (2000) 339–350, https://doi.org/10.2118/65705-PA.
[34] L.B. Ouyang, K. Aziz, A mechanistic model for gas-liquid flow in horizontal wells with radial influx or outflux, Petrol. Sci. Technol. 20 (1–2) (2002) 191–222, https://doi.org/10.1081/LFT-120002095.
[35] J.J. Xiao, O. Shoham, J.P. Brill, Comprehensive mechanistic model for two-phase flow in pipelines, in: 65th SPE Annual Technical Conference and Exhibition, New Orleans, LA, September 23–26, 1990, 1990, pp. 167–180, https://doi.org/10.2118/20631-MS.
[36] A.R. Hasan, C.S. Kabir, A study of multiphase flow behavior in vertical wells, SPE Prod. Eng. 3 (2) (1988a) 263–272, https://doi.org/10.2118/15138-pa.
[37] A.R. Hasan, C.S. Kabir, Predicting multiphase flow behavior in a deviated well, SPE Prod. Eng. 3 (4) (1988b) 474–482, https://doi.org/10.2118/15449-PA.
[38] P.M. Ozon, G. Ferschneider, A. Chwetzoff, A new multiphase flow model predicts pressure and temperature profiles in wells, SPE Offshore Eur. 87 (1987) 8–11, https://doi.org/10.2118/16535-ms. *Aberdeen*.
[39] N. Petalas, K. Aziz, Mechanistic model for multiphase flow in pipes, J. Can. Petrol. Technol. 39 (6) (2000) 43–55, https://doi.org/10.2118/00-06-04.
[40] H.Q. Zhang, C. Sarica, Unified modeling of gas/Oil/Water-Pipe flow-basic approaches and preliminary validation, SPE Proj. Facil. Constr. 1 (2) (2006) 1–7, https://doi.org/10.2118/95749-PA.
[41] H.Q. Zhang, Q. Wang, C. Sarica, J.P. Brill, Unified model for gas-liquid pipe flow via slug dynamics - Part 1: model development, J. Energy Resour. Technol. 125 (58) (2003) 266–273, https://doi.org/10.1115/1.1615246.
[42] I. Jahanandish, B. Salimifard, H. Jalalifar, Predicting bottomhole pressure in vertical multiphase flowing wells using artificial neural networks, J. Petrol. Sci. Eng. 75 (3–4) (2011) 336–342, https://doi.org/10.1016/j.petrol.2010.11.019.
[43] G. Takacs, Considerations on the selection of an optimum vertical multiphase pressure drop prediction model for oil wells, in: Proceedings - SPE Production Operations Symposium, 2001, pp. 663–672, https://doi.org/10.2118/68361-ms.
[44] C.C. Nwanwe, U.I. Duru, Comparison and performance analysis of models for predicting multiphase flow behaviours in wellbores, Int. J. Petrol. Geosci. Eng. 2022 (2022) 1–20.
[45] E.A. Osman, M.A. Ayoub, M.A. Aggour, Artificial Neural Network Model for Predicting Bottomhole Flowing Pressure in Vertical Multiphase Flow, in: SPE Middle East Oil and Gas Show and Conference, 2005, pp. 1–11, https://doi.org/10.2118/93632-ms.
[46] X. Li, J.L. Miskimins, B.T. Hoffman, A combined bottom-hole pressure calculation procedure using multiphase correlations and artificial neural network models, in: SPE Annual Technical Conference and Exhibition, 1–12, 2014, https://doi.org/10.2118/170683-ms.
[47] C.C. Nwanwe, U.I. Duru, C. Anyadiegwu, A.I.B. Ekejuba, An artificial neural network visible mathematical model for real-time prediction of multiphase flowing bottom-hole pressure in wellbores, Petrol. Res. xxxx (2022), https://doi.org/10.1016/j.ptlrs.2022.10.004.
[48] N.A. Sami, D.S. Ibrahim, Forecasting multiphase flowing bottom-hole pressure of vertical oil wells using three machine learning techniques, Petrol. Res. xxxx (2021), https://doi.org/10.1016/j.ptlrs.2021.05.004.
[49] D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey, Artificial Intelligence Review (2018), https://doi.org/10.1007/s10462-017-9610-2.
[50] F. Agin, R. Khosravanian, M. Karimifard, A. Jahanshahi, Application of Adaptive Neuro-Fuzzy Inference System and Data Mining Approach to Predict Lost Circulation Using DOE Technique (Case Study: Maroon Oilfield), Petroleum 6 (4) (2020) 423–437, https://doi.org/10.1016/j.petlm.2018.07.005.
[51] P. Chen, M.M. Rahman, H.K. Sarma, A. Dhabi, Interaction between hydraulic fracture and natural fracture - a new prediction model by adaptive neuro-fuzzy inference system (ANFIS), in: Abu Dhabi International Petroleum Exhibition and Conference, 2014, https://doi.org/10.2118/171927-ms.
[52] M.A. Ahmadi, Z. Chen, Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs, Petroleum 5 (3) (2019) 271–284, https://doi.org/10.1016/j.petlm.2018.06.002.
[53] M. Babanezhad, I. Behroyan, A.T. Nakhjiri, A. Marjani, S. Shirazian, Performance and application analysis of ANFIS artificial intelligence for pressure prediction of nanofluid convective flow in a heated pipe, Sci. Rep. (2021) 1–18, https://doi.org/10.1038/s41598-020-79628-w.
[54] H. Yavari, R. Khosravanian, D.A. Wood, B.S. Aadnoy, Application of mathematical and machine learning models to predict differential pressure of autonomous downhole inflow control devices, Adv. Geo-Energy Res. 5 (4) (2021) 386–406, https://doi.org/10.46690/ager.2021.04.05.
[55] F.S. Alakbari, M.E. Mohyaldinn, M.A. Ayoub, A.S. Muhsan, I.A. Hussein, A reservoir bubble point pressure prediction model using the Adaptive Neuro-Fuzzy Inference System (ANFIS) technique with trend analysis, PLoS One 17 (8 August) (2022) 1–19, https://doi.org/10.1371/journal.pone.0272790.
[56] A. Al-Shammari, Prediction of Pressure Drop for Two-phase Flow in Vertical Pipes Using Artificial Intelligence, M.S. Thesis. King Fahd University of Petroleum and Minerals, 2011b.
[57] R. Moazenzadeh, B. Mohammadi, M.J.S. Safari, K. wing Chau, Soil moisture estimation using novel bio-inspired soft computing approaches, Eng. Appl. Comput. Fluid Mech. 16 (1) (2022) 826–840, https://doi.org/10.1080/19942060.2022.2037467.
[58] S. Mehdizadeh, B. Mohammadi, F. Ahmadi, Establishing coupled models for estimating daily dew point temperature using nature-inspired optimization algorithms, Hydrology 9 (1) (2022), https://doi.org/10.3390/hydrology9010009.
[59] S. Mehdizadeh, B. Mohammadi, Q.B. Pham, Z. Duan, Development of boosted machine learning models for estimating daily reference evapotranspiration and comparison with empirical approaches, Water (Switzerland) 13 (24) (2021), https://doi.org/10.3390/w13243489.
[60] F. Hadadi, R. Moazenzadeh, B. Mohammadi, Estimation of actual evapotranspiration-A novel hybrid method based on remote sensing and artificial intelligence.pdf, J. Hydrol. 609 (2022), 127774, https://doi.org/10.1016/j.jhydrol.2022.127774.
[61] S. Emamgholizadeh, A. Bazoobandi, B. Mohammadi, H. Ghorbani, Prediction of soil cation exchange capacity using enhanced machine learning approaches in the southern region of the caspian sea, Ain Shams Eng. J. xxxx (2022), 101876, https://doi.org/10.1016/j.asej.2022.101876.
[62] Z. Tariq, M. Mahmoud, A. Abdulraheem, Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques, J. Pet. Explor. Prod. Technol. 10 (4) (2020) 1411–1428, https://doi.org/10.1007/s13202-019-0728-4.

[63] M.A. Ayoub, Development and Testing of an Artificial Neural Network Model for Predicting Bottomhole Pressure in Vertical Multiphase Flow. M.S. Thesis, King Fahd University of Petroleum and Minerals, 2004.
[64] M.A. Ahmadi, Z. Chen, Machine learning models to predict bottom hole pressure in multi-phase flow in vertical oil production wells, Can. J. Chem. Eng. (2019) 1–13, https://doi.org/10.1002/cjce.23526.
[65] O. Akinsete, B.A. Adesiji, Bottom-hole pressure estimation from wellhead data using artificial neural network, in: SPE Nigeria Annual International Conference and Exhibition, 2019, https://doi.org/10.2118/198762-ms.
[66] M.A. El-Kadi, M.A. Husien, S.M. El-Behery, H. Farouk, Prediction of two-phase pressure drop using artificial neural network, Eng. Res. J 42 (2) (2019) 99–114.
[67] E.A. Kanin, A.L. Vainshtein, A.A. Osiptsov, E.V. Burnaev, The method of calculation the pressure gradient in multiphase flow in the pipe segment based on the machine learning algorithms, IOP Conf. Ser. Earth Environ. Sci. 193 (2018), https://doi.org/10.1088/1755-1315/193/1/012028.
[68] Y. Seong, C. Park, J. Choi, I. Jang, Surrogate model with a deep neural network to evaluate gas – liquid flow in a horizontal pipe, Energies 13 (4) (2020) 1–12, https://doi.org/10.3390/en13040968.
[69] J.R. Jang, Anfis : adaptive-network-based fuzzy inference system, in: IEEE Transactions on Systems, Man, and Cybernetics, 1993, https://doi.org/10.1109/21.256541, 23(3.
[70] N. Talpur, M.N.M. Salleh, K. Hussain, An investigation of membership functions on performance of ANFIS for solving classification problems, IOP Conf. Ser. Mater. Sci. Eng. 226 (1) (2017), https://doi.org/10.1088/1757-899X/226/1/012103.
[71] S. Aminifar, G. Yosefi, Application of adaptive neuro fuzzy inference system (ANFIS) in implementing of new CMOS fuzzy logic controller (FLC) chip, AIP Conf. Proc. 936 (2007) 49–53, https://doi.org/10.1063/1.2790190.
[72] Z. Liao, F. Zhang, Identification of nonlinear system based on ANFIS with Hybrid fuzzy clustering, Inf. Technol. J. 12 (24) (2013) 8349–8353, https://doi.org/10.3923/itj.2013.8349.8353.
[73] K. Premkumar, B.V. Manikandan, Fuzzy PID supervised online ANFIS based speed controller for brushless dc motor, Neurocomputing 157 (2015) 76–90, https://doi.org/10.1016/j.neucom.2015.01.032.
[74] H. Monirvaghefi, M. Rafiee Sandgani, M. Aliyari Shoorehdeli, Interval type-2 adaptive network-based fuzzy inference system (ANFIS) with type-2 non-singleton Fuzzification, in: 13th Iranian Conference on Fuzzy Systems, IFSC 2013, 2013, https://doi.org/10.1109/IFSC.2013.6675612.
[75] P. Hájek, V. Olej, Intuitionistic fuzzy neural network: the case of credit scoring using text information, Commun. Comput. Inf. Sci 517 (2015) 337–346, https://doi.org/10.1007/978-3-319-23983-5_31.
[76] A. Dehghanian Serej, H. Mojallali, Speed control of elliptec motor using adaptive neural-fuzzy controller with on-line learning simulated under MATLAB/SIMULINK, in: IntelliSys 2015 - Proceedings of 2015 SAI Intelligent Systems Conference, 2015, pp. 543–553, https://doi.org/10.1109/IntelliSys.2015.7361193.
[77] M. Turki, S. Bouzaida, A. Sakly, F. M'Sahli, Adaptive control of nonlinear system using neuro-fuzzy learning by PSO algorithm, in: Proceedings of the Mediterranean Electrotechnical Conference - MELECON, 2012, pp. 519–523, https://doi.org/10.1109/MELCON.2012.6196486.
[78] J.J. Cárdenas, A. García, J.L. Romeral, K. Kampouropoulos, Evolutive ANFIS training for energy load profile forecast for an IEMS in an automated factory, in: IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, 1–8, 2011, https://doi.org/10.1109/ETFA.2011.6059079.
[79] D. Karaboga, E. Kaya, Training ANFIS using artificial bee colony algorithm, in: 2013 IEEE International Symposium on Innovations in Intelligent Systems and Applications, IEEE INISTA 2013, 1–5, 2013, https://doi.org/10.1109/INISTA.2013.6577625.
[80] H.N. Nhu, S. Nitsuwat, M. Sodanil, Prediction of stock price using an adaptive Neuro-fuzzy inference system trained by firefly algorithm, in: 2013 International Computer Science and Engineering Conference, ICSEC 2013, 2013, pp. 302–307, https://doi.org/10.1109/ICSEC.2013.6694798.
[81] K. Hussain, M.N.M. Salleh, Optimization of fuzzy neural network using APSO for predicting strength of Malaysian SMEs, in: 2015 10th Asian Control Conference: Emerging Control Techniques for a Sustainable World, ASCC 2015, 2015, https://doi.org/10.1109/ASCC.2015.7244638.
[82] O.F. Lutfy, S.B. Mohd Noor, M.H. Marhaban, A simplified adaptive neuro-fuzzy inference system (ANFIS) controller trained by genetic algorithm to control nonlinear multi-input multi-output systems, Sci. Res. Essays 6 (31) (2011) 6475–6486, https://doi.org/10.5897/sre11.1059.
[83] D. Karaboga, E. Kaya, An adaptive and hybrid artificial bee colony algorithm (aABC) for ANFIS training, Appl. Soft Comput. J. 49 (2016) 423–436, https://doi.org/10.1016/j.asoc.2016.07.039.
[84] J. Brownlee, Data preparation for machine learning-data cleaning, feature selection and data transforms in Python, Machine Learning Mastery, 2020.
[85] B. Illowsky, S. Dean, Introductory. Stat., Issue 9 53, OpenStax, 2018.
[86] O. Shoham, Mechanistic Modeling of Gas Liquid Two Phase Flow in Pipes, SPE Textbook Series, Richardson, TX, 2006.
[87] L. Márquez-Torres, J. Ochoa-Pineda, P. Pico, J.P. Valdés, D. Becerra, A. Pinilla, E. Pereyra, N. Ratkovich, Comparison of 63 different void fraction correlations for different flow patterns, pipe inclinations, and liquid viscosities, SN Appl. Sci. 2 (10) (2020) 1–24, https://doi.org/10.1007/s42452-020-03464-w.