



Development of a new closed-form symbolic regression correlation for predicting multiphase flow wellhead pressure through chokes: A case study of Buzurgan oil field

Mohamed H. Mohamed ^{a,*}, Murtadha K. Abdali ^a, Baqer W. Falh ^a, Hassan A. Ibrahim ^a

a Department of Petroleum Engineering, College of Oil and Gas Engineering, University of Technology, Baghdad, Iraq

Abstract

Accurate calculation and monitoring of multiphase choke-flow pressure drops present a critical engineering challenge for optimizing production rates and ensuring effective reservoir management. Traditional empirical correlations, such as Gilbert and Ros formulations, often show reduced accuracy when applied to fields like Buzurgan in Iraq. Although modern machine learning models offer high statistical accuracy, their “black-box” nature limits practical field deployment due to the lack of transparent, explicit formulas. To bridge this gap, this study introduces an artificial intelligence framework using Python symbolic regression (PySR) to develop a novel, transparent, and directly deployable closed-form algebraic correlation using 672 field datasets from the Buzurgan Oil Field based on choke size, liquid flow rate, and gas-liquid ratio. The derived equation achieved an R^2 of approximately 0.80, an RMSE of 162.6 psi, and a MAPE of 18.1% on test data, outperforming traditional correlations ($R^2 = 0.60$ – 0.65). Comparisons with random forest, DNN, XGBoost, and ANN showed that while random forest achieved the highest accuracy ($R^2 = 0.94$), the PySR model offers a practical compromise between accuracy and interpretability, providing a transparent engineering trade-off that sacrifices minor statistical precision for an explicit formula essential for immediate field deployment and daily operational optimization. The proposed explicit algebraic formula enables production engineers to implement the correlation directly in spreadsheets and field software, improving daily production optimization. This study contributes to solving the “black box” problem by providing a powerful, accurate, and straightforward tool tailored to Iraqi oil field conditions.

Keywords: Wellhead Choke; Multiphase Flow; Wellhead Pressure; Symbolic Regression; Buzurgan Oil Field.

Received on 12/06/2026, Received in Revised Form on 01/08/2026, Accepted on 01/08/2026, Published on 30/09/2026

<https://doi.org/10.31699/IJCPE.2026.3.6>

1- Introduction

The wellhead chokes represent an important element in surface oil and gas production facilities, known as the production tree (Christmas tree). Such tools are used primarily for regulation of the flow rate, maintaining stability of the wellbore, managing the pressure drop at reservoirs, and mitigating negative effects like water or gas coning and sanding. Therefore, calculation of multiphase flow properties and pressure drop through the chokes is a very important aspect for proper optimization and efficient management of the reservoir [1, 2].

Traditionally, production engineers have been using conventional correlations for evaluation of the performance of multiphase choke, such as correlations suggested by Gilbert [3] and Ros [4]. Although the advantage of using these conventional closed-form equations is that they are simple to use and easily applied in the field, the disadvantage of such correlations was that they were traditionally fitted based on certain local datasets. Therefore, such correlations demonstrate poor performance and high prediction errors when applied to

other geological and hydrodynamic conditions, like, for instance, oil fields in Iraq (like Buzurgan field). These specific correlations were selected due to their historical significance and widespread adoption in production engineering literature as benchmarks for choke performance.

Eq. 1 represents the basic mathematical relationship introduced by Gilbert [3]:

$$P_{wh} = \frac{C \cdot q_1 \cdot GLR^m}{S^n} \quad (1)$$

Table 1 shows the variables used for each correlation [5, 6].

In an effort to overcome the inherent nonlinearities of multiphase flow through choke valves, modern research has widely utilized sophisticated techniques from Artificial Intelligence (AI) and Machine Learning (ML) fields to achieve greater statistical accuracy in the predictions of choke performance [10, 11]. For instance, Khamis et al. [10] evaluated four distinct AI techniques including Artificial Neural Networks (ANN), Fuzzy



*Corresponding Author: Email: mohamed.haider8090@gmail.com

© 2026 The Author(s). Published by College of Engineering, University of Baghdad.

This is an Open Access article licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). This permits users to copy, redistribute, remix, transmit and adapt the work provided the original work and source is appropriately cited.

Logic (FL), Support Vector Machines (SVM), and Functional Networks (FN) to optimize choke sizes and predict flow behavior, demonstrating that advanced computational structures can effectively map complex wellhead operations. Similarly, Akbari and Rahimi [11] employed a hybrid Multi-Layer Perceptron integrated with a Genetic Algorithm (MLP-GA) model, trained on an optimized dataset of 875 clean points (filtered from an initial 910 samples), to forecast well performance parameters with an outstanding testing R^2 of 0.9936. Despite the integration of post-hoc interpretability

methods such as SHapley Additive exPlanation (SHAP) into these frameworks in order to evaluate feature importance and determine input impacts such as choke dimensions and fluid rates on pressure predictions [12] it is crucial to recognize that SHAP is merely an ex-post explanation tool applied after model training. It does not alter the underlying black-box architecture; thus, traditional machine learning algorithms inherently retain significant “black-box” limitations for daily field utilization [13, 14].

Table 1. Empirical constants for correlations

Correlation	C	m	n
Gilbert [3]	10.0	0.546	1.89
Ros [4]	17.40	0.500	2.00
Baxendell [7]	9.56	0.546	1.93
Beiranvand et al. [8]	30.49	0.589	2.275
Okon et al. [9]	5.1474	0.5048	1.7093

Therefore, a fundamental distinction must be drawn between post-hoc Explainable AI (XAI) tools like SHAP which merely interpret complex models after training and symbolic regression, which directly generates transparent, explicit equations. Through the embedding of the physics into thousands of hidden layers and weights of the model, these algorithms become practically useless for daily optimizations in the field without specialized software, a limitation that fundamental white-box frameworks aim to resolve. While previous AI implementations (such as SVM, FL, and MLP-GA) captured complex non-linear behaviors, they either functioned as black-box structures or relied on ex-post explanations without providing closed-form equations. Moreover, these models were largely validated on general datasets rather than field-specific data subjected to harsh thermodynamic conditions like those in Iraqi oil fields.

Furthermore, standard empirical correlations (such as Gilbert and Ros) were originally derived under simplified fluid dynamics that fail to account for the complex reservoir and fluid behaviors observed in Iraqi oil fields [15, 16]. For instance, in the Buzurgan Oil Field (southern Iraq), the primary producing formation (Mishrif) exhibits strong petrophysical heterogeneities in porosity and water saturation across key production units [17]. More critically, recent fluid thermodynamic evaluation of live-oil samples from the Buzurgan field reveals complex hydrocarbon behavior characterized by high risk of asphaltene precipitation and deposition under pressure drop conditions [18, 19]. Sudden pressure drops occurring across wellhead chokes alter the fluid's phase behavior and multiphase flow characteristics. Because classical empirical correlations do not incorporate these localized thermodynamic and petrophysical phenomena, they yield significant prediction errors, necessitating the development of a novel, physics-informed correlation calibrated explicitly for Iraqi operational conditions [20]. To bridge this crucial gap, the current paper presents a unique method based on Symbolic Regression (SR) using

the PySR framework that acts as an effective tool connecting the predictive accuracy of multidimensional AI algorithms and the applicability of traditional closed-form equations [21]. PySR was specifically selected over alternative symbolic regression algorithms due to its advanced multi-threaded execution in Julia and its superior ability to optimize the Pareto frontier, effectively balancing mathematical complexity against predictive accuracy [22].

To address these persistent limitations, this study bridges the existing research gap by introducing a new, highly accurate, and explicit closed-form artificial intelligence correlation specifically optimized for predicting multiphase flow pressure drop through chokes in Iraqi oil fields. The primary objective of this work is to leverage the robust mathematical search capabilities of symbolic regression via PySR to formulate a transparent equation that eliminates the “black-box” barrier. The key novelty lies in developing a physics-informed, easily deployable mathematical expression that matches the statistical accuracy of complex neural networks while remaining directly applicable for daily, routine field optimization by production engineers without the need for specialized computational software.

2- Methodology

The evaluation of the best method depends on its capabilities, practicality, and real-world applicability. In this study, older models, which were developments of Eq. 1, were evaluated alongside modern black-box models and a linear regression model. Fig. 1 illustrates the procedure used in the practical aspect of this research to evaluate the best method among the proposed approaches.

• Dataset and benchmark models

The primary database used in this study consists of historical and daily field records of "production tests,"

obtained directly from the Buzurgan oil field for several wells such as BU3, BU4, BU5, BU6, BU9, BU10, and BU11. These raw files include detailed information on wellhead performance. Initially, the raw data contained a very wide range of operational parameters and variables, including: choke size (d), oil production rate (Q_o) associated water ratio characteristics (Q_w) and gas production flow rate (Q_g), gas-to-oil ratio (GOR), sediment-water ratio (BSW), flow pressure at wellhead (P_{wh}), as well as manifold (collector) pressure, buffer

pressure, salinity ratio, and fluid density. After this, the data was filtered to extract only the input relevant to our study, and rows containing missing data were discarded to avoid hindering training and evaluation, Table 2 shows some of the samples of the filtered data. The total number of data rows is 1023, but after removing missing data, zero values, and converting choke sizes (d) greater than 1.0 into normalized fractional inches (d/64), the final dataset was reduced to 672 clean rows.

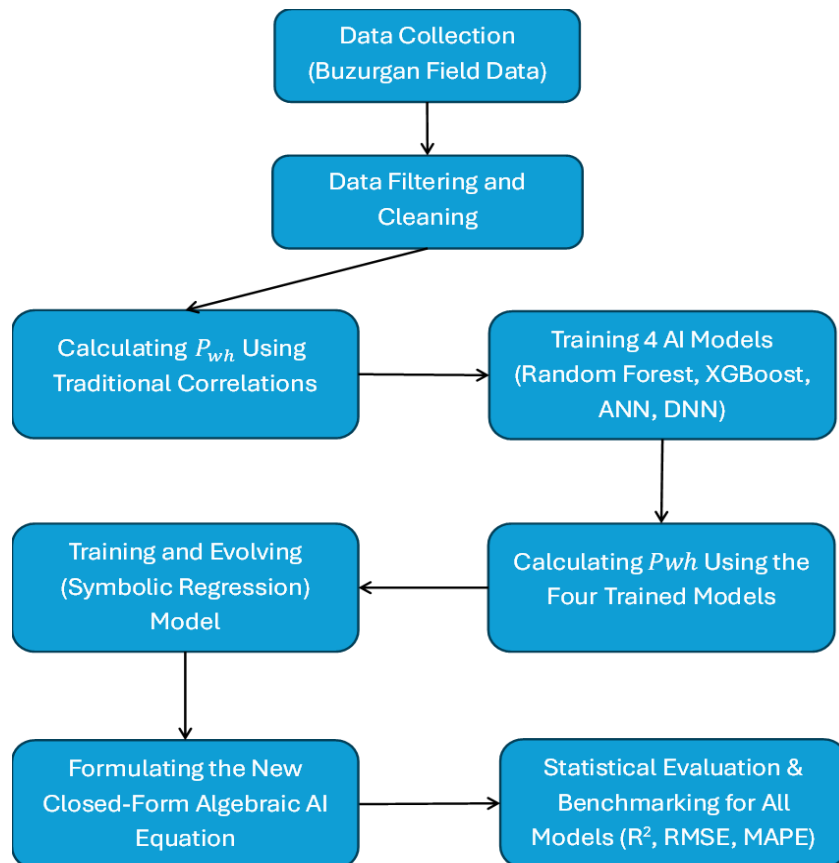


Fig. 1. Flowchart of the methodology used to create a new equation and evaluate it against other models

Table 2. Samples of data after filtering

CHOKE fraction	Size(d)	CHOKE Size(d) $s = x/64$	Liquid Flow Rate (m^3/day)	GLR (SCF/STB)	Actual Pwh (psi)
0.5625	36/64		767.5	654.66	1166.31
0.5625	36/64		207	457.03	369.81
0.625	40/64		668	562.02	1152.09
0.5	32/64		427	516.54	1066.75
0.75	48/64		752	641.75	924.51

Table 2 shows some of the samples of the filtered data. Input data in preprocessing stage were normalized in order to make them relative to the black box algorithms (such as DNN, ANN) in order to ensure stable gradient descent and prevent feature dominance. However, in the case of the PySR algorithm, there is no need for

normalization or scaling, because symbolic regression models do not depend on the scale of the input data [23]. The data was divided into a training period of 75% and a testing period of 25% randomly to ensure that the models did not retain data and to test them on a large dataset to determine their R^2 .

- Training AI models

For a more general view and to make a comparison, some of the AI models like random forest, DNN, XGBoost, and ANN were trained. It helped in making a more thorough assessment since it allowed observing the R^2 match for each of them and comparing their differences. As far as the black box models go, they are

mostly used in the code base rather than being written as explicit formulas; hence, the applicability of the black box models is usually less compared to explicit formulas that can be produced by the symbolic regression model, for instance. The structural flowchart and specific hidden layer architectures for the utilized AI models are detailed in Fig. 2.

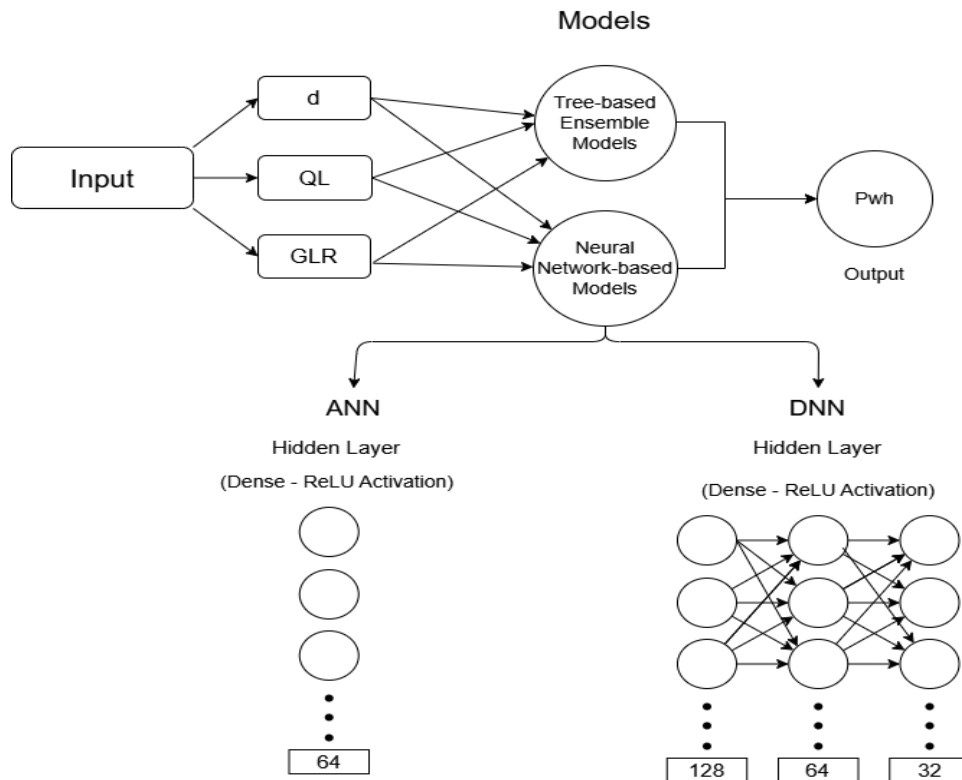


Fig. 2. Structural flowchart of the deployed tree-based and neural network models, including specific hidden layer architectures for ANN and DNN configurations

- Evolutionary AI framework (PySR)

After identifying this problem, we used an advanced generation of artificial intelligence that "creates" new equations from scratch based on the evolutionary engineering of the Buzurgan oil field data. The data we have consists of inputs such as [liquid flow rate (m^3/day), GLR (SCF/STB), and choke size as percentage], which are the same inputs for the Gilbert equation, and the output was the wellhead pressure (Pwh) in (psi) unit.

The symbolic regression (PySR) engine powered by Julia was adopted, by utilizing the open-source PySR framework, the search algorithm efficiently navigates complex mathematical landscapes using a highly parallelized symbolic architecture built on Julia. Unlike generic neural architectures that pre-determine structure and only optimize weights, this evolutionary approach treats mathematical equations as dynamic expressions, optimizing both the functional form and the constant parameters simultaneously to identify analytical laws from physical datasets [24] which searches the

mathematical space for the best programming constructs that connect variables kinematically and hydraulically.

Also, to prevent the generation of overly complex black-box formulations and maintain high physical interpretability, the evolutionary programming engine was strictly restricted to the operational function set and structural search bounds detailed in Table 3.

The performances of the empirical correlations used in the past were compared to that of the proposed closed form equation using the coefficient of determination (R^2), which is shown in Eq. 2 [25]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

Where y_i represents the actual measured field wellhead pressure P_{wh} and $\hat{y}_i$ denotes the predicted pressure value calculated by the developed AI closed-form correlation also $\bar{y}$ is the arithmetic mean of the actual production dataset and n is the total number of tested (field/production data). Additionally, the final expression was selected through a rigorous optimization process where multiple candidate equations were evaluated. The

optimal equation was ultimately chosen based on the mathematical simplicity and high predictive accuracy (R²).
pareto frontier to ensure an ideal balance between

Table 3. PySR model configuration and search space constraints

Parameter / Feature	Setting / Selection	Implementation Details
Data Partitioning	Train 75% and Test 25%	Applied on 672 cleaned records (random_state = 42)
Operator Space	Binary: +, -, ×, /, ^ Unary: 1/x, sin, cos, tan	Covers kinematic & hydraulic non-linear terms
Architectural Limits	Max power = 5 Blocked nested trig (sin(sin x))	Prevents structural overfitting and oscillation
Selection Criterion	Pareto Frontier (R ² Max on 25% Test set)	Mean Squared Error (MSE) over 300 iterations

3- Results and discussion

The results, as shown in Table 4, strongly suggest Random Forest, DNN (Deep Neural Network), XGBoost, and symbolic regression (PySR) are the best, even in the testing phase, meaning they didn't store data and didn't

fall into the overfitting trap. Next came the equation produced by the symbolic regression model, which represents a middle ground, or even a stronger one, as it combines the power of artificial intelligence with ease of practical application. Following this in the rankings came Gilbert original equation with its modifications.

Table 4. Statistical performance comparison of machine learning models and empirical correlations

Model	RMSE (test) (psi)	MAPE (test)%	R ² (train)	R ² (test)
Random Forest	98.9	7.30	0.98	0.94
DNN	112.4	13.0	0.93	0.89
XGBoost	120.8	7.49	0.99	0.87
Symbolic Regression (PySR)	162.6	18.1	0.81	0.80
ANN	170.7	26.70	0.74	0.74
Okon et al. [9]	232.6	20.7	-	0.65
Gilbert [3]	328.5	28.6	-	0.65
Baxendell [7]	237.7	17.8	-	0.65
Ros [4]	251.5	16.8	-	0.64
Beiranvand et al. [8]	334.4	26.1	-	0.60

The XGBoost model initially appeared to be the best model, but it was quickly tested and its accuracy decreased, with R² being 0.99 in training and 0.87 in testing. This indicates that the model fell into the overfitting trap, where it memorized rather than understood. Despite this, it still showed good accuracy in testing. We also note that the symbolic regression (PySR) model is a good model compared to the other correlations, as it has an explicit equation Eq. 3 and also average to good accuracy. It combines the power of neural networks with the ease of writing the equation. It is a white-box model that can produce an explicit equation that can be used later directly without needing to use an AI model stored in the computer's memory. That is, although the Random Forest model achieved R² = 0.94 in testing, which is numerically higher than PySR, which achieved R² = 0.80, Random Forest remains a blind (black-box) model that a field engineer cannot write into a quick equation in Excel. Herein lies the sacrifice of PySR (trade-off), as it sacrifices some statistical accuracy in exchange for this closed equation (Eq. 3).

$$p_{wh}(PySR) = -434.50867 \cos\left(d \times \frac{6895.4785}{Q_L + 458.22284}\right) + 750.9685 \quad (3)$$

Where p_{wh} in (psi), d in percentage, Q_L in (m³/day).

For Eq. 3, the internal value contains Q_L, which means, geometrically speaking, that when the choke diameter is constant and the fluid flow rate (Q_L) increases, the overall value of the cosine angle decreases. Since the cosine function increases as the angle approaches zero (in the first quadrant), multiplying it by a negative sign -434.5 will ultimately result in a pressure drop or change that is hydraulically proportional to the flow and choke. Physically, this cosine function acts as a mathematical bounding mechanism that captures the transitional behavior and pressure fluctuations across different multiphase flow regimes (such as slug or bubble flow) inside the choke.

The number 458.22284 added to Q_L in the denominator is a very clever mathematical safety factor (damping/regularization factor) that prevents the algorithm from falling into the trap of dividing by zero when the well is shut down or the flow stops Q_L = 0.

Furthermore, the omission of the Gas-Liquid Ratio (GLR) by the PySR algorithm suggests that within the operating envelopes and fluid properties of the studied Iraqi reservoirs, the total pressure drop is hydrodynamically dominated by liquid-phase momentum and constriction friction losses, making gas volume variations secondary for a simplified closed-form expression. Therefore, when comparing the (PySR) equation with traditional equations the PySR equation (Eq. 3) appears as the closest equation to the true values of p_{wh} , even without the GLR in the equation, since the symbolic regression model sacrificed the small positive effect of the GLR in order to formulate a simple equation that is neither long nor complex.

The best model in terms of accuracy is Random Forest, which achieved excellent accuracy in training data. This is attributed to its strong structure for generalization

outside the training data, as its accuracy did not decrease significantly between training and testing. From a fluid mechanics perspective, this superiority occurs because tree-based ensemble algorithms efficiently map the highly non-linear and chaotic hydrodynamic interactions of multiphase flow across chokes without assuming rigid, pre-defined parametric boundaries, aligning with recent comparative studies in choke modeling literature. However, in terms of clarity and ease of application, the symbolic regression (PySR) model is superior to other correlations such as Gilbert and its modifications, as seen in Fig. 3. The proposed new equation exhibits the highest accuracy among the other equations when calculating P_{wh} . Therefore, it can also calculate the liquid flow rate more accurately than other methods when it is unavailable.

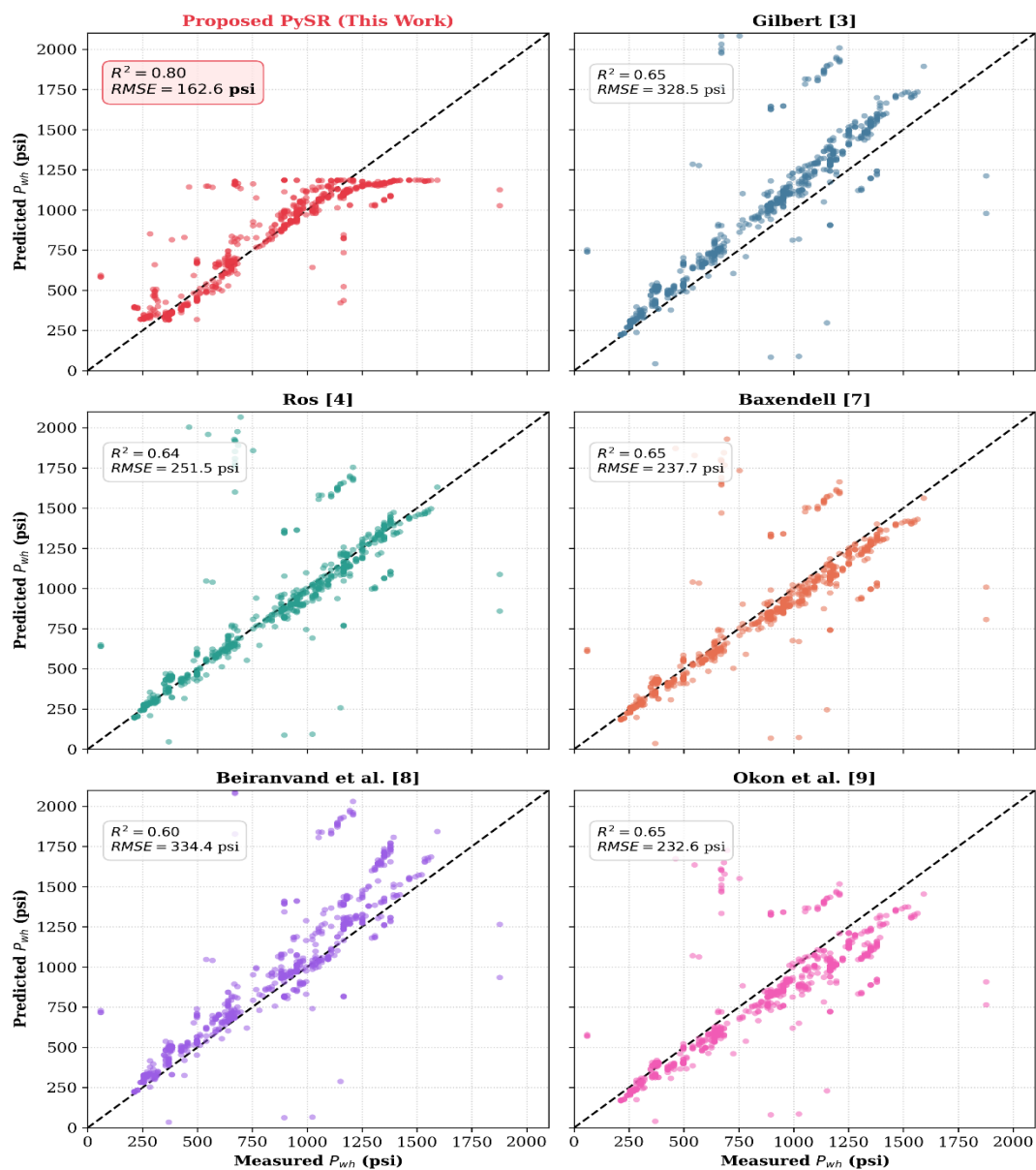


Fig. 3. Parity plots measured versus predicted wellhead pressure (P_{wh}) for all evaluated models

4- Conclusions

This study successfully introduced a novel, explicit closed-form white-box correlation derived via Julia-powered symbolic regression (PySR) to predict multiphase flow pressure drops across wellhead chokes in the Buzurgan Oil Field (southern Iraq). The resulting symbolic correlation (Eq. 3) demonstrated strong predictive accuracy ($R^2 = 0.80$, RMSE = 162.7 psi) on unseen field testing data, significantly outperforming standard empirical correlations (such as Gilbert, Ros, Baxendell, Beiranvand, and Okon), which exhibited poor predictive stability with R^2 values ranging from 0.60 to 0.65 under Iraqi operational conditions. Although black-box machine learning algorithms (e.g., Random Forest) achieved higher purely statistical metrics ($R^2 = 0.94$), model "optimality" in field production engineering is defined by an objective multi-criteria trade-off rather than raw numerical accuracy alone.

This evaluation framework balances predictive accuracy, mathematical transparency, computational independence, and daily operational deployability. PySR represents the optimal solution within this multi-criteria space because it yields an explicit algebraic equation that petroleum engineers can directly implement in standard spreadsheets and field monitoring tools without relying on complex software dependencies or specialized computational environments. Despite its strong practical utility, several key limitations of this work must be explicitly acknowledged. The proposed correlation was calibrated exclusively using production records from the Buzurgan Oil Field (Mishrif formation), meaning that applying this equation outside its calibrated thermodynamic and fluid property ranges may require recalibration. Furthermore, the model has not yet undergone external validation across geographically distinct reservoirs, nor does it incorporate formal uncertainty or sensitivity analyses. Future research should focus on testing the generalizability of this symbolic framework across diverse Middle Eastern oil fields, incorporating formal uncertainty quantification, and embedding physics-informed boundary conditions directly into the evolutionary search function.

Nomenclature

Symbol / Abbreviation	Description and Units
Symbol	
c, m, n	Empirical constants used in Gilbert and Ros correlations (dimensionless)
d	Choke size (inches)
Q_L	Liquid flow rate (m ³ /day)
P_{wh}	Wellhead pressure, measured or predicted (psi)
Q_o	Oil production flow rate (m ³ /day)
GLR	Gas-Liquid Ratio (SCF/STB)
R^2	Coefficient of determination (dimensionless)

y_i Actual measured field wellhead pressure value (psi)

Abbreviations

AI	Artificial Intelligence
MAPE	Mean Absolute Percentage Error
PySR	Python Symbolic Regression
RMSE	Root Mean Squared Error
SR	Symbolic Regression

References

- [1] B. Guo, W. C. Lyons, and A. Galiullin, *Petroleum Production Engineering: A Computer-Assisted Approach*. Amsterdam: Elsevier Science & Technology Books, 2007, pp. 59-67. <https://doi.org/10.1016/B978-0-7506-8270-1.X5000-2>
- [2] M. J. Economides, A. D. Hill, C. Ehlig-Economides, and D. Zhu, *Petroleum Production Systems*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2013, pp. 236-247.
- [3] W. E. Gilbert, "Flowing and gas-lift well performance," *API Drilling and Production Practice*, vol. 1954, pp. 126-157, Jan. 1954.
- [4] N. C. J. Ros, "An analysis of critical simultaneous gas-liquid flow through a restriction," *Applied Scientific Research*, Section A, vol. 9, no. 5, pp. 374-388, 1960. <https://doi.org/10.1007/BF00382215>
- [5] A. N. Okon and D. Appah, "Neural Network Models for Predicting Wellhead Pressure-Flow Rate Relationship for Niger Delta Oil Wells," *Journal of Scientific Research and Reports*, vol. 12, no. 1, pp. 1-14, Jan. 2016. <https://doi.org/10.9734/JSRR/2016/28715>
- [6] A. Lak, R. Azin, S. Osfouri, and R. Fatehi, "Modeling Critical Flow through Choke for a Gas-condensate Reservoir Based on Drill Stem Test Data," *Iranian Journal of Oil & Gas Science and Technology*, vol. 6, no. 3, pp. 29-40, 2017. <https://doi.org/10.22050/ijogst.2017.69063.1371>
- [7] P. B. Baxendell, "Producing Wells on Casing Flow – An Analysis of Flowing Pressure Gradients," *Transactions of the AIME*, vol. 213, pp. 202-206, 1958. <https://doi.org/10.2118/983-G>
- [8] M. S. Beiranvand, P. Mohammadmoradi, B. Aminshahidy, B. Fazelabdolabadi, and S. Aghahoseini, "New multiphase choke correlations for a high flow rate Iranian oil field," *Mechanical Sciences*, vol. 3, no. 1, pp. 43-47, Jun. 2012. <https://doi.org/10.5194/ms-3-43-2012>
- [9] A. N. Okon, F. D. Udoh, and D. Appah, "Empirical Wellhead Pressure-Production Rate Correlations for Niger Delta Oil Wells," *presented at the SPE Nigeria Annual International Conference and Exhibition*, Lagos, Nigeria, Aug. 2015. <https://doi.org/10.2118/178303-MS>

- [10] M. Khamis, M. Elhaj, and A. Abdulraheem, "Optimization of choke size for two-phase flow using artificial intelligence," *Journal of Petroleum Exploration and Production Technology*, vol. 10, no. 2, pp. 487–500, May 2020. <https://doi.org/10.1007/s13202-019-0734-6>
- [11] A. Akbari and M. Rahimi, "Prediction of well performance parameters for wellhead choke flow optimization using MLP-GA," *Geosystem Engineering*, 2026. <https://doi.org/10.1080/12269328.2026.2613005>
- [12] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [13] W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowledge-Based Systems*, vol. 263, p. 110273, Mar. 2023. <https://doi.org/10.1016/j.knosys.2023.110273>
- [14] I. U. Ekanayake, D. P. P. Meddage, and U. Rathnayake, "A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)," *Case Studies in Construction Materials*, vol. 16, e01059, 2022. <https://doi.org/10.1016/j.cscm.2022.e01059>
- [15] A. Akbari, F. Ghazi, and Y. Kazemzadeh, "Machine learning-based prediction of well performance parameters for wellhead choke flow optimization," *Scientific Reports*, vol. 15, art. 44182, Dec. 2025. <https://doi.org/10.1038/s41598-025-27851-8>
- [16] F. S. Alakbari, M. A. Ayoub, M. A. Awad, T. Ganat, M. E. Mohyaldinn, and S. M. Mahmood, "A robust pressure drop prediction model in vertical multiphase flow: a machine learning approach," *Scientific Reports*, vol. 15, art. 13420, Apr. 2025. <https://doi.org/10.1038/s41598-025-96371-2>
- [17] M. M. Mohammed, H. M. Salih, and K. H. Mnaty, "3D Reservoir Modeling of Buzurgan Oil Field, Southern Iraq," *Iraqi Journal of Science*, vol. 63, no. 2, pp. 596–607, Feb. 2022. <https://doi.org/10.24996/ij.s.2022.63.2.16>
- [18] M. A. Ahmed, J. M. Al Said Naji, A. Ali, and A. Qassim Al-Khafaje, "Asphaltene Phase Behavior Modeling Using Peng-Robinson 78A-EOS for a Crude Oil Sample from Buzurgan Oil Field," *Petroleum Chemistry*, vol. 64, no. 7, pp. 840–848, Oct. 2024. <https://doi.org/10.1134/S096554412405013X>
- [19] B. Amiri-Ramsheh, S. Sahebalzamani, R. Zabihi, and A. Hemmati-Sarapardeh, "Predicting asphaltene precipitation during natural depletion of oil reservoirs by integrating SARA fractions with advanced intelligent models," *Scientific Reports*, vol. 15, art. 26214, Jul. 2025. <https://doi.org/10.1038/s41598-025-11966-z>
- [20] M. S. Dabiri, F. Hadavimoghaddam, S. Ashoorian, M. Schaffie, and A. Hemmati-Sarapardeh, "Modeling liquid rate through wellhead chokes using machine learning techniques," *Scientific Reports*, vol. 14, no. 1, art. 6945, Mar. 2024. <https://doi.org/10.1038/s41598-024-54010-2>
- [21] L. S. Keren, A. Liberzon, and T. Lazebnik, "A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge," *Scientific Reports*, vol. 13, art. 1249, Jan. 2023. <https://doi.org/10.1038/s41598-023-28328-2>
- [22] A. Tonda, "Review of PySR: high-performance symbolic regression in Python and Julia," *Genetic Programming and Evolvable Machines*, vol. 26, art. 7, Dec. 2024. <https://doi.org/10.1007/s10710-024-09503-4>
- [23] S. M. Koksang, "Cosmic backreaction and the mean redshift drift from symbolic regression," *Phys. Rev. D*, vol. 107, p. 103522, May 2023. <https://doi.org/10.1103/PhysRevD.107.103522>
- [24] Udrescu, S. M., and Tegmark, M., "AI Feynman: A physics-inspired method for symbolic regression," *Science Advances*, vol. 6, no. 16, p. eaay2631, Apr. 2020. <https://doi.org/10.1126/sciadv.aay2631>
- [25] Chicco, D., Warrens, M. J., and Jurman, G., "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. e623, Jul. 2021. <https://doi.org/10.7717/peerj-cs.623>

تطوير علاقة ارتباطية جديدة للانحدار الرمزي بصيغة مغلقة للتنبؤ بضغط رأس البئر لتدفق متعدد الأطوار عبر الخوانق: دراسة حالة لحقل بزركان النفطي

محمد حيدر محمد^{1*}، مرتضى كاظم عبد علي¹، باقر واثق فالح¹، حسن علي ابراهيم¹

¹ قسم هندسة النفط، كلية هندسة النفط والغاز، الجامعة التكنولوجية، بغداد، العراق

الخلاصة

يمثل الحساب الدقيق ومراقبة انخفاضات الضغط في تدفق الاختناق متعدد الأطوار تحدياً هندسياً بالغ الأهمية لتحسين معدلات الإنتاج وضمان الإدارة الفعالة للمكان. غالباً ما تُظهر العلاقات التجريبية التقليدية، مثل صيغ جيلبرت وروس، دقة منخفضة عند تطبيقها على حقول مثل حقول بزروران في العراق. على الرغم من أن نماذج التعلم الآلي الحديثة توفر دقة إحصائية عالية، إلا أن طبيعتها المبهمة تحد من استخدامها العملي في الحقول نظراً لافتقارها إلى صيغ واضحة وشفافة. لسد هذه الفجوة، تقدم هذه الدراسة إطار عمل للذكاء الاصطناعي باستخدام الانحدار الرمزي في بايثون (PySR) لتطوير علاقة جبرية مغلقة جديدة وشفافة وقابلة للتطبيق مباشرة، وذلك باستخدام ٦٧٢ مجموعة بيانات حقلية من حقول بزروران النفطي، بناءً على حجم الاختناق ومعدل تدفق السائل ونسبة الغاز إلى السائل. حققت المعادلة المشتقة معامل تحديد (R^2) يقارب ٠,٨٠، ومتوسط الجذر التربيعي للخطأ (RMSE) 162.6 رطل/بوصة مربعة، ومتوسط النسبة المئوية للخطأ المطلق 18.1% (MAPE) على بيانات الاختبار، متفوقاً بذلك على الارتباطات التقليدية ($R^2 = 0.60$). وأظهرت المقارنات مع نماذج الغابة العشوائية، والشبكات العصبية العميقة (DNN)، وXGBoost، والشبكات العصبية الاصطناعية (ANN) أنه على الرغم من أن نموذج الغابة العشوائية حقق أعلى دقة ($R^2 = 0.94$)، إلا أن نموذج PySR يقدم حلاً وسطاً عملياً بين الدقة وسهولة التفسير، موفراً بذلك حلاً هندسياً واضحاً يضيء بدقة إحصائية طفيفة مقابل صيغة صريحة ضرورية للتطبيق الميداني الفوري وتحسين العمليات اليومية. تُمكن الصيغة الجبرية الصريحة المقترحة مهندسي الإنتاج من تطبيق الارتباط مباشرة في جداول البيانات وبرامج الحقول، مما يُحسن من كفاءة الإنتاج اليومية. تُساهم هذه الدراسة في حل مشكلة "الصندوق الأسود" من خلال توفير أداة قوية ودقيقة ومباشرة مصممة خصيصاً لظروف حقول النفط العراقية.

الكلمات الدالة: خناق رأس البئر، التدفق متعدد الأطوار، ضغط رأس البئر، الانحدار الرمزي، حقل نفط البزركان.