

A COMPARATIVE ANALYSIS OF THE PREDICTION OF GAS CONDENSATE DEW POINT PRESSURE USING MACHINE LEARNING ALGORITHMS AND TRADITIONAL EMPIRICAL CORRELATIONS

Alejandro Osorio Pozo¹, Lucio Álvarez Pérez¹ , Karim Botan², Timur Vasile Chis³ 

¹ Universidad Veracruzana, Facultad de Ciencias Químicas, Coatzacoalcos, México.

² Doctoral School, Engineering Sciences (Mines, Oil and Gas), Petroleum-Gas University of Ploiesti, Romania.

³ Petroleum-Gas University of Ploiesti, Romania.

* e-mail: alosorio@uv.mx; lucalvarez@uv.mx

DOI: 10.51865/JPGT.2026.01.21

ABSTRACT

Accurate prediction of dew point pressure (P_d) is critical for managing gas condensate reservoirs, as liquid dropout near the wellbore creates condensate banking that reduces permeability and well productivity. This study proposes a data-driven approach using a Random Forest (RF) machine learning algorithm to predict P_d based on reservoir temperature and fluid composition. Utilizing a comprehensive dataset of 375 records, the model incorporates 13 predictors, including hydrocarbon fractions (C_1 through C_{7+}), non-hydrocarbons (N_2 , CO_2 , H_2S), and heavy fraction properties (MC_{7+} , γ_{C7+}).

The proposed RF model exhibits highly competitive accuracy, outperforming most traditional empirical correlations by achieving a superior overall data variance capture with a Coefficient of Determination (R^2) of 0.8790 and an Average Absolute Percent Relative Error (APE) of 8.17%. While the Ahmadi-Elsharkawy model shows a marginally lower APE (7.90%), the RF framework avoids complex genetic programming equations and delivers superior global consistency across the entire pressure envelope. By capturing complex, non-linear thermodynamic interactions, the RF model provides a robust, fast, and cost-effective alternative to expensive laboratory PVT tests and complex equations of state, optimizing fluid characterization and production system design.

Keywords: Reservoir temperature, dew point pressure, gas composition, molecular weight, specific gravity, retrograde condensation, gas condensate, heavy components, Random Forest, Machine Learning.

INTRODUCTION

In gas condensate reservoirs, well productivity often declines sharply when near-wellbore pressure falls below the dew point. Literature suggests that radial compositional-reservoir models are effective for investigating this phenomenon, as they demonstrate how liquid dropout creates a ring of high condensate saturation around the wellbore [1-3]. This condensate banking reduces the effective permeability to gas, leading to a rapid loss in productivity.

Consequently, accurately determining the dew point pressure is vital for effective reservoir management [4],[5]. Experimental determination of dew point pressure typically involves

laboratory PVT tests. These include Constant Mass Expansion (*CME*), conducted in a visual window-type cell, and Constant Volume Depletion (*CVD*), which verifies thermodynamic equilibrium and tracks compositional changes as pressure decreases [6],[7].

For predictive modeling, Equations of State (*EOS*) can be tuned to match experimental data using various binary interaction parameters. Other methods rely on K-value correlations, though these often require tedious trial-and-error computations and lose accuracy at high temperatures and pressures. Ultimately, precise dew point prediction is essential for fluid characterization, production system design, and avoiding unnecessary simulation overhead. While laboratory measurements remain the gold standard for reliability, technical or economic constraints often necessitate the use of these robust modeling and correlation alternatives [7],[8].

In the context of predicting dew point pressure in gas condensate systems, decision tree-based models have shown outstanding performance in capturing complex interactions between gas composition, temperature, and heavy fraction properties, yielding superior results compared to those obtained through equations of state or conventional empirical correlations [8].

Recent studies have reinforced these findings by demonstrating that advanced machine learning models, such as XGBoost, significantly outperform empirical estimators in terms of accuracy, adaptability, and robustness, even with limited databases and high compositional variability [9]. Traditional empirical correlations often lose accuracy when applied to fluid systems that differ from their original development datasets.

This study addresses these limitations by implementing a Random Forest (RF) machine learning algorithm, which transitions from rigid mathematical formulas to a data-driven, soft computing approach [10]. The RF algorithm was selected for its ability to capture complex, non-linear interactions between reservoir temperature, gas composition, and heavy fraction properties, because the model constructs multiple decision trees on random subsets of data and averages their outputs. In this averaging mechanism ensures the model is robust and generalizes well to new, unseen reservoir data and our model was developed using the Scikit-learn library in Python, utilizing 13 specific predictors including C_1 through C_{7+} , N_2 , CO_2 , H_2S , and C_{7+} physical properties [11].

LITERATURE SURVEY

Determining dew point pressure is essential for managing gas condensate fluids. Beyond complex mathematical models and equations of state, several widely recognized empirical correlations offer a practical alternative for predicting this critical parameter.

Nemeth and Kennedy correlation calculates dew point pressure based on the mole fractions of C_1 through C_{7+} , CO_2 , H_2S , and N_2 , along with the physical properties (molecular weight and specific gravity) of the C_{7+} fraction [2]. The authors developed this eleven-constant equation from a using 579 data points from 489 distinct gas condensate systems, their model is applicable within a pressure range of 1270–10790 psia and a temperature range of 40–320°F.

Elsharkawy proposed an equation for dew point pressure based on 340 data points (131 experimental and 209 from literature) [12]. This model is applicable across a wide range of conditions, covering pressures from 1560 to 11830 psia, temperatures from 40 to 340°F, and gas-condensate gravities between 0.65 and 1.89.

Notably, the model accounts for both sweet and highly corrosive sour gases, as well as the full range of fluid richness (from lean to rich C_{7+} content).

Shokir in 2008 developed a mathematical model for dew point pressure using Genetic Programming, based on a dataset of 245 experimental gas condensate samples [13]. Although the model covers a wide operational range of temperatures and gas properties, it excludes the specific gravity of the C_{7+} fraction [13].

Ahmadi and Elsharkawy correlation is a robust, and intelligent model used to predict the dew point pressure of gas condensate reservoirs, generally outperforming traditional empirical methods [14]. It is developed using gene expression programming (GEP), providing high accuracy by utilizing compositional data (C_1 – C_{7+}), molecular weight and specific gravity of the C_{7+} fraction and reservoir temperature (T_R). The model covers a pressure range of 1400–10790 psia and a temperature from 76 to 320°F. The data bank contains some sour gas condensate that has as high as 30% H_2S and others having 92% CO_2 . It also covers a very wide range of gas condensate compositions; some samples are very lean and have methane as high as 91% and low heptane plus, others are very rich and have a heptane content as high as 13%.

Nnadozie in 2018 proposed a predictive model for the dew point pressure of gas condensate reservoirs based on reservoir temperature and routinely available gas compositional data. Developed using a comprehensive dataset of 259 gas condensate samples sourced from specialized literature, the model accounts for a wide range of fluid properties and thermal conditions [7].

DATA ACQUISITION

The accuracy and reliability of any predictive model depend fundamentally on the quality and sufficiency of the input data. Consistent with recent research, dew-point pressure (P_d) in gas condensate reservoirs is modelled as a function of reservoir temperature, fluid composition (both hydrocarbon and non-hydrocarbon), and the specific characteristics of the C_{7+} fraction [14–18]. In this study, the model was developed and validated using a dataset of 375 points, consisting of 75 experimental measurements and 300 values sourced from literature [19–23]. This dataset covers a broad experimental envelope, including temperatures from 40°F to 313°F, C_{7+} specific gravities from 0.730 to 0.850, and C_{7+} molecular weights from 106 to 253. The corresponding dew-point pressures range from 2405 psia to 11830 psia, as detailed in Table 1.

MODEL DEVELOPMENT AND IMPLEMENTATION

Due to the complex relationship between fluid composition and dew-point pressure, developing a universal global correlation remains a significant challenge. Furthermore, experimental determination of dew-point pressure for gas condensate fluids is both costly and time-consuming. This study establishes a robust mathematical relationship between experimental data and key reservoir outputs using a soft computing approach (Table 1).

Algorithm Selection and Hyperparameter Tuning

The Random Forest algorithm, implemented via the Scikit-learn library in Python, was selected for its ability to model complex non-linear relationships and its inherent resistance to overfitting. The model was optimized using grid search cross-validation, establishing the following key hyperparameters to ensure stability: $n_estimators = 100$, $max_depth = 15$, and $random_state = 42$.

Table 1. Range and statistical parameters of the used data for model development [10]

Parameter	Minimum	Average	Maximum	Standard Deviation
P_d (Psia)	2405	4954.12	11830	1552.15
T_R (°F)	40	217.02	313	50.56
M_{C7+}	106	152.22	253	24.03
γ_{C7+}	0.7300	0.7906	0.8500	0.0236
x_{C1}	0.3344	0.7886	0.9650	0.1174
x_{C2}	0.0040	0.0610	0.1453	0.0303
x_{C3}	0.0023	0.0314	0.0902	0.0191
$x_{C4's}$	0.0022	0.0199	0.0630	0.0118
$x_{C5's}$	0.0009	0.0127	0.1230	0.0132
$x_{C6's}$	0.0000	0.0095	0.0871	0.0094
x_{C7+}	0.0036	0.0427	0.1356	0.0314
x_{N2}	0.0000	0.0128	0.4322	0.0408
x_{CO2}	0.0000	0.0135	0.1036	0.0171
x_{H2S}	0.0000	0.0068	0.2657	0.0317

Mathematical Description of the Model

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the mean prediction of the individual trees for regression tasks.

Mathematically, given a training dataset

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, \quad (1)$$

In the relationship 1 we defined:

- D represents the entire training dataset used to build the Random Forest model.
- n represents the total number of experimental data points (or samples) available in the dataset (in the context of your study, $n = 375$).
- i is the index representing a specific data point, ranging from 1 to n .
- x_i : Denotes the input feature vector for the i -th sample. Mathematically, it is expressed as $x_i \in \mathbb{R}^{13}$, meaning it belongs to a 13-dimensional real coordinate space because it contains exactly 13 independent predictor variables.

Specifically, this vector includes:

- Reservoir temperature (T_R).
- Hydrocarbon molar fractions (x_{C1} through x_{C7+}).
- Non-hydrocarbon molar fractions (x_{N2} , x_{CO2} , x_{H2S}).
- Physical properties of the heavy fraction (molecular weight M_{C7+} and specific gravity γ_{C7+}).
- y_i denotes the target output variable corresponding to the input vector x_i . Mathematically, it is expressed as $y_i \in \mathbb{R}$ (a single real number) and represents the actual, experimentally measured dew point pressure (P_d) for that specific fluid sample.

For each individual decision tree $t = 1, 2, \dots, T$ (where $T = n_{estimators} = 100$):

1. Draw a bootstrap sample D_t of size n randomly with replacement from the total training dataset D .
2. Grow a regression tree $h_t(x)$ using the bootstrapped data D_t by recursively repeating the following steps for each terminal node of the tree until the specified constraints are satisfied ($max_depth = 15$):
 - Select m variables at random from the 13 available predictors.
 - Determine the best variable and split-point among the m variables by minimizing the node impurity. For regression, this split criterion is dictated by minimizing the Mean Squared Error (MSE) within a given node region R_m :

$$MSE = \frac{1}{|R_m|} \sum_{i \in R_m} (y_i - \bar{y}_m)^2 \quad (2)$$

where $\bar{y}_m$ is the mean of the target values within the local region R_m .

- Split the node into two distinct daughter nodes based on the optimal split-point.

Once the entire ensemble of T decision trees is fully trained, the final predicted dew point pressure $\hat{y}$ for any new input vector x is generated by aggregating and averaging the outputs of all individual regression trees:

$$\hat{y}(x) = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (3)$$

This bootstrap aggregating (bagging) mechanism significantly decreases the overall model variance without increasing its bias, leading to excellent generalization capabilities and robust prevention of overfitting on unseen reservoir fluid configurations.

Input Variable Selection

The architecture utilizes 13 specific predictors to estimate dew-point pressure. As illustrated in Figure 1, these inputs include reservoir temperature (T_R), molar fractions of hydrocarbons (C_1 to C_{7+}), non-hydrocarbons (N_2 , CO_2 , H_2S), and the physical properties of the heavy fraction (M_{C7+} and γ_{C7+}).

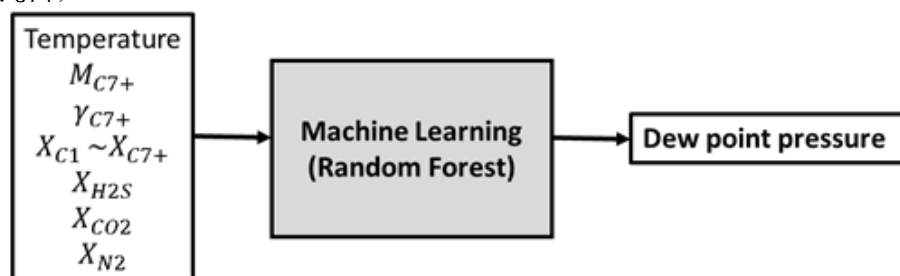


Figure 1. Developed RF model scheme for prediction of dew point pressure.

Training and Performance Evaluation

To ensure robust generalization, the total dataset of 375 records was divided using a standard 80/20 train-test split, supplemented by k -fold cross-validation during the training phase. The model's validity was further assessed against 36 independent external datasets taken from Elsharkawy (Table 2). Comparison between the measured/predicted results of the developed model and the applied data of dew-point pressure for Sample 1 and Sample 2 are presented in Figure 2 and Figure 3 respectively.

Table 2. Comparison between the predicted and target data.

Sample set	Dew point pressure (psia)		Error (%)
	Measurement	Predicted (RF)	
1	3095	2982.72	3.63
2	3337	5468.91	63.89
3	2651	2675.05	0.91
4	11830	8586.32	27.42
5	3345	3104.45	7.19
6	8750	8517.52	2.66
7	5780	5199.08	10.05
8	5229	6311.05	20.69
9	4203	4822.89	14.75
10	4173	4825.52	15.64
11	5219	5263.58	0.85
12	4172	4748.27	13.81
13	4160	4907.68	17.97
14	7871	6868.73	12.73
15	3838	4279.01	11.5
16	3413	4856.69	42.31
17	4076	4235.33	3.92
18	4854	5083.53	4.72
		APE	15.26

Sample set	Dew point pressure (psia)		Error (%)
	Measurement	Predicted (RF)	
1	8540	8419.58	1.41
2	5720	5438.32	4.92
3	6668	7076.34	6.12
4	10790	8046.31	25.43
5	9655	8541.95	11.53
6	4372	5640.23	29.01
7	1405	1723.07	22.64
8	4423	3169.37	28.34
9	8500	8838.83	3.99
10	1835	2420.04	31.88
11	3385	3405.76	0.61
12	3445	3867.34	12.26
13	5870	6032.15	2.76
14	3520	3287.56	6.6
15	4630	4689.8	1.29
16	4414	4311.1	2.33
17	3915	4160.85	6.28
18	4465	4363.19	2.28
		APE	9.63

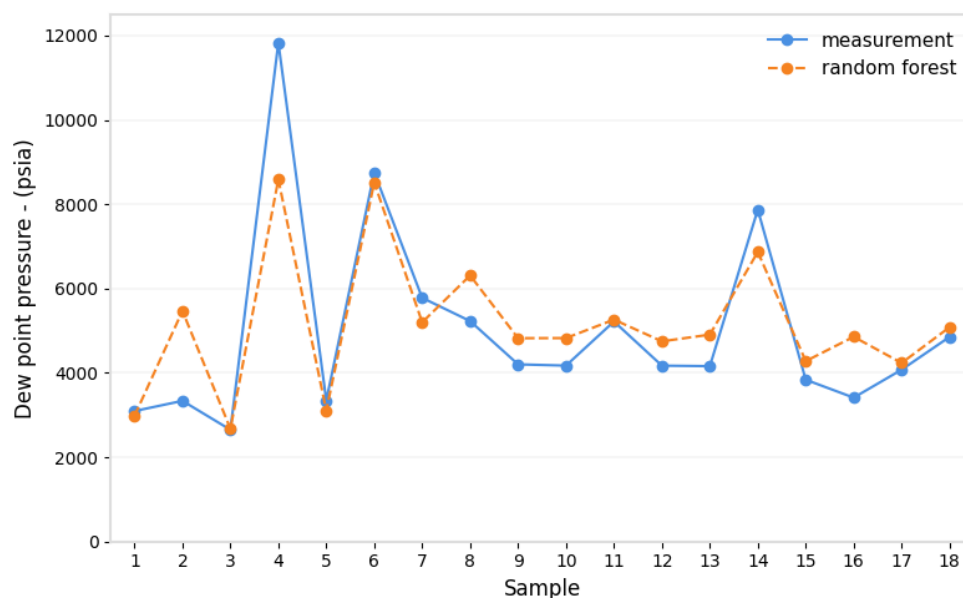


Figure 2. Comparison between the measured/predicted results of the developed model and the applied data of dew-point pressure - Sample 1.

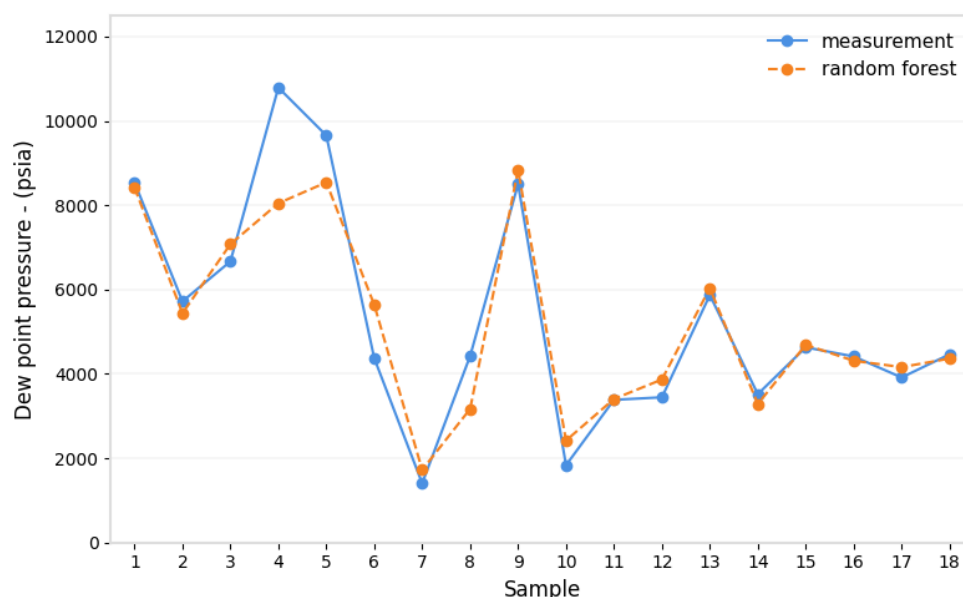


Figure 3. Comparison between the measured/predicted results of the developed model and the applied data of dew-point pressure – Sample 2.

Model performance was evaluated using multiple fitness functions, including the Coefficient of Determination (R^2), Average Percent Relative Error (APE), Root Mean Square Error ($RMSE$), and Mean Absolute Error (MAE).

While the overall model demonstrates high accuracy, specific outliers in Sample 1 (e.g., records 2 and 16) exhibit larger relative errors; these deviations correlate with extreme compositional limits and highly sour gas mixtures, indicating areas where future model calibration could be focused.

RESULTS AND DISCUSSION

Dew point pressure correlations determined by Nemeth-Kennedy, Elsharkawy, Shokir, Ahmadi-Elsharkawy, and Nnadozie were used for comparison against the new model. Due to the significant variation in gas-condensate fluid compositions across different global reservoirs, existing correlations often yield inaccurate predictions when applied to fluid systems outside their original development datasets. A comparative analysis and graphical error assessment are detailed in Figure 4 and Table 3. The results demonstrate that the proposed RF -model maintains a tight distribution along the identity line. While the Ahmadi-Elsharkawy correlation achieved a slightly lower APE (7.90%) compared to the RF model (8.17%), the RF approach exhibited superior overall variance capture. Specifically, the RF model achieved a significantly higher R^2 of 0.8790 compared to Ahmadi-Elsharkawy's 0.7765. This indicates that the RF model reproduces experimental behavior with higher global precision and consistency across the entire pressure range of 2405 to 11830 psia, avoiding the extreme complexities associated with genetic programming equations [24],[25].

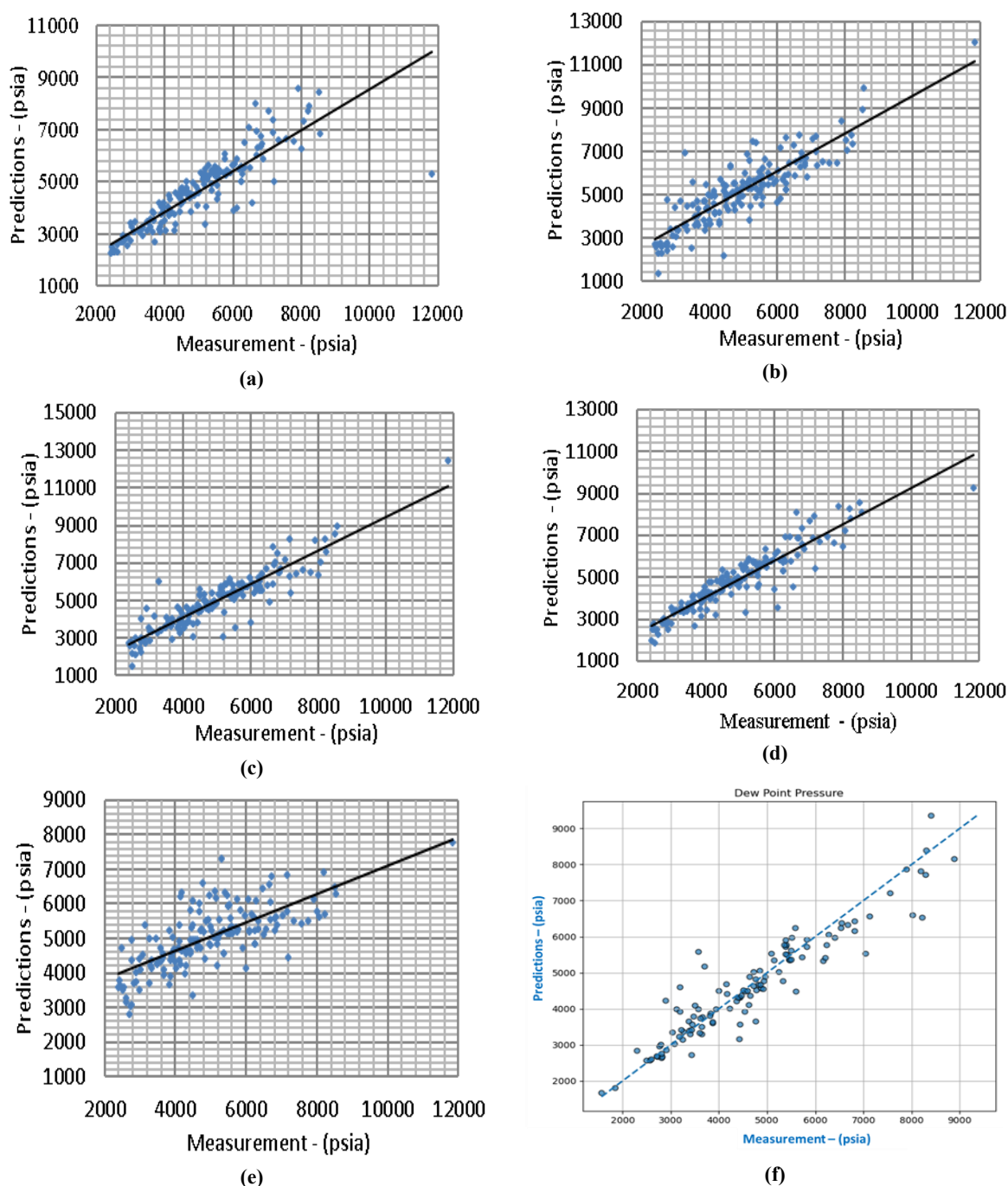


Figure 4. Cross plot of measured versus predicted dew point pressure using (a) Nemeth-Kennedy, (b) Elsharkawy, (c) Shokir, (d) Ahmadi-Elsharkawy, (e) Nnadozie, (f) this study.

Table 3. Statistical analysis results of the correlations used in this comparative study.

Quantity	Nemeth-Kennedy	Elsharkawy	Shokir	Ahmadi-Elsharkawy	Nnadozie	This study
APE (%)	8.69	13.54	9.71	7.90	18.87	8.17
R-square	0.7397	0.5937	0.8250	0.7765	0.5985	0.8790

In our experiments:

RMSE (psia) is 983.99,

MAE (psia) is 648.44,

Max Error (psia) is 3243.68.

1. Root Mean Square Error (*RMSE*):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

This equation measures the standard deviation of the prediction errors. Squaring the errors before averaging heavily penalizes larger errors/outliers.

2. Mean Absolute Error (*MAE*):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

This equation calculates the average magnitude of the errors in a set of predictions, without considering their direction.

3. Maximum Error (*Max Error*):

$$MaxError = \max(|y_i - \hat{y}_i|) \quad (6)$$

This equation simply identifies the single largest absolute difference between a predicted value and its corresponding experimental value.

4. Average Percent Relative Error (*APE*):

$$APE = 1/n \sum |y_i - \hat{y}_i| / y_i \times 100 \quad (7)$$

This equation expresses the absolute error as a percentage of the actual measured value, providing a clear normalized metric of how far off the predictions are on average.

In equation 4-7 we defined:

- n represents the total number of data points (or samples) being evaluated in your dataset.
- i represents the specific index of a given data point, iterating from 1 up to n .
- y_i denotes the actual (target) experimentally measured value of the dew point pressure for the i -th sample.
- $\hat{y}_i$ denotes the predicted value of the dew point pressure generated by your model (or correlation) for the i -th sample.
- terms $\sum$ (or $\sum_{i=1}^n$) the summation symbol, indicating that you must add up the calculated values for all n samples in the dataset.
- $|\dots|$ denotes the absolute value. This operation converts any negative numbers to positive numbers, ensuring that over-predictions and under-predictions do not cancel each other out when added together.
- $\max$ denotes the maximum value. In equation (6), it means finding the single largest absolute difference between a predicted value and its corresponding experimental value.

These terms are standard across machine learning and statistical modeling for evaluating how close a model's predictions are to the real-world data.

Model Interpretability and Error Diagnostics

To provide deeper insights into the model's decision-making process and move beyond global metrics, a Feature Importance analysis was conducted. By extracting the *feature_importances_attribute* from the trained Scikit-Learn Random Forest model, the analysis reveals that reservoir temperature (TR) and the physical properties of the heavy fraction (MC_{7+} and γ_{C7+}) are the dominant drivers in predicting the dew point pressure, followed by the molar fractions of heavy components (C_6 , C_5), as illustrated in Figure 5. This confirms the model successfully learned core thermodynamic behaviors.

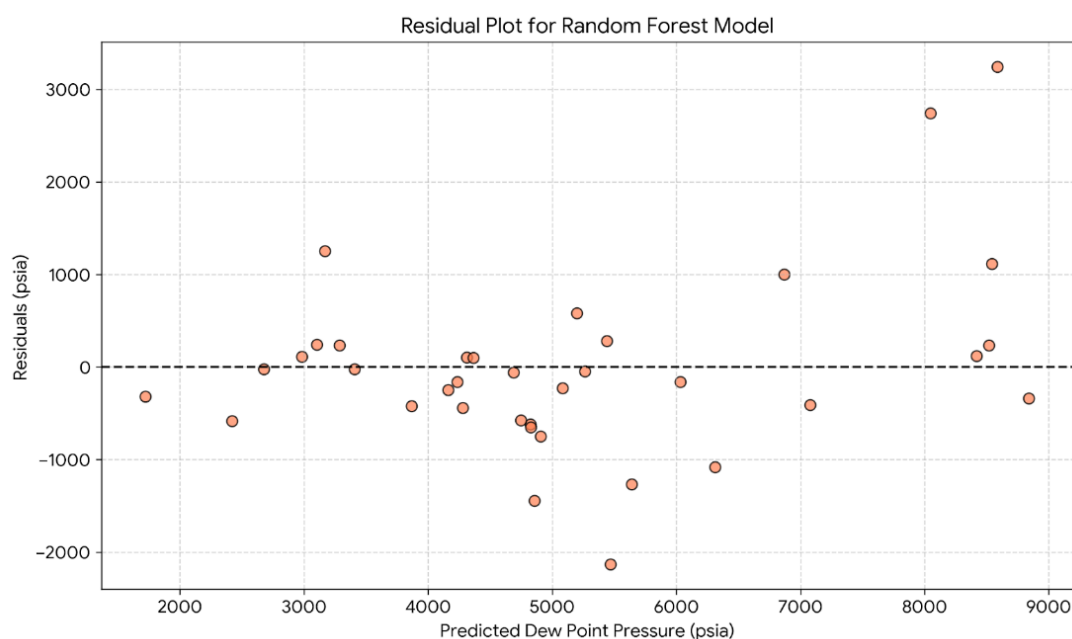


Figure 5. Feature importance of the 13 predictors used in the Random Forest model for dew point pressure estimation.

Furthermore, to evaluate error distribution rigorously, a Residual Plot (Figure 6) was generated (calculated as $Residuals = Actual P_d - Predicted P_d$). Visualizing the residuals on the Y-axis against the predicted pressure on the X-axis (Figure 6) demonstrates that the prediction errors are randomly scattered around the zero line ($Y=0$) across the entire operational envelope of 2405 to 11830 psia. This confirms that the RF model does not suffer from systematic bias (it does not consistently over-predict or under-predict) in any specific pressure range.

CONCLUSIONS

This study transitioned from traditional rigid mathematical formulas to a data-driven Machine Learning approach to predict dew point pressure (P_d) in gas condensate reservoirs. Based on the experimental and statistical results, the data-driven Random Forest (RF) model provides a highly effective and robust alternative to traditional empirical correlations. The developed RF architecture achieved a strong Coefficient of Determination (R^2) of 0.8790 and an Average Percent Relative Error (APE) of 8.17%.

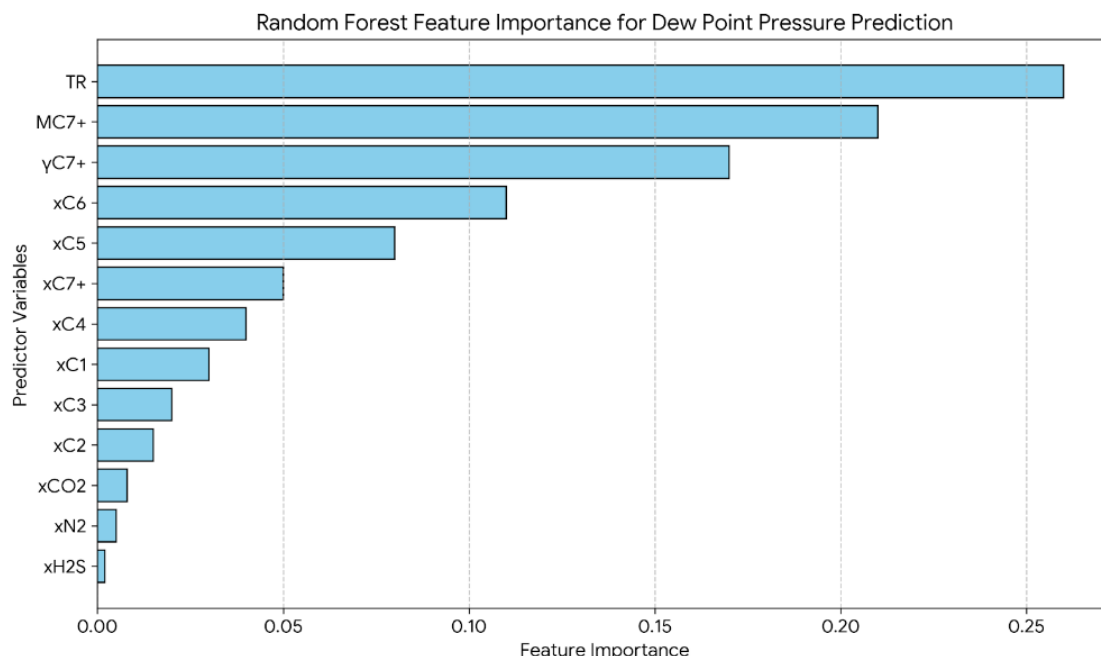


Figure 6. Residual plot showing the distribution of prediction errors against predicted dew point pressure.

While the Ahmadi-Elsharkawy model yielded a slightly lower *APE* (7.90%), our proposed *RF* framework captures the global data variance with visibly greater precision, as evidenced by the superior R^2 metric (0.8790 vs. 0.7765). More importantly, the machine learning approach completely avoids the rigid and highly complex mathematical formulations characteristic of genetic programming equations. This ensures a more adaptable, straightforward, and computationally efficient implementation while delivering robust predictive precision across a remarkably broad operational pressure envelope ranging from 2405 to 11830 psia.

From an operational and economic perspective, the implementation of this soft computing technique offers a fast, reliable, and cost-effective alternative for reservoir engineering workflows. The data-driven framework successfully bypasses the need for time-consuming and expensive laboratory PVT testing protocols, such as Constant Mass Expansion (*CME*) and Constant Volume Depletion (*CVD*). Furthermore, it eliminates the necessity for tedious, repetitive trial-and-error computations traditionally associated with legacy K-value correlations.

In conclusion, from a practical reservoir management standpoint, the precise estimation of dew point pressure via this Random Forest model provides production engineers with an invaluable tool to anticipate and mitigate near-wellbore condensate banking. Understanding the exact pressure thresholds at which liquid dropout occurs allows operators to design and execute proactive pressure maintenance strategies. Ultimately, this prevents localized permeability reduction around the wellbore, sustaining and optimizing long-term well productivity and economic recovery.

Nomenclature

M_{C7+} – Molecular weight of C_{7+} fraction.	X_{C1} – Molar fraction of methane.
γ_{C7+} – Specific gravity of C_{7+} fraction.	X_{C2} – Molar fraction of ethane.
P_d – Dew point pressure, psia.	X_{C3} – Molar fraction of propane.
T_R – Reservoir temperature, °F.	X_{C4} – Molar fraction of butane.
X_{N2} – Molar fraction of nitrogen.	X_{C5} – Molar fraction of pentane.
X_{CO2} – Molar fraction of carbon dioxide.	X_{C6} – Molar fraction of hexane.
X_{H2S} – Molar fraction of hydrogen sulfide.	X_{C7+} – Molar fraction of heptane's plus.

Appendix A. Equations for the comparative models used in this study

This study compares the results of the new machine learning based approach with traditional models from Nemeth-Kennedy, Elsharkawy, Shokir, Ahmadi-Elsharkawy and Nnadozie. The following equations represent the standard correlations used in this comparative assessment:

Nemeth and Kennedy (1967)

Table A.1. The coefficients of Nemeth and Kennedy correlation [10].

$A_1 = -2.0623054$	$A_2 = 6.6259728$	$A_3 = -4.4670559 \cdot 10^{-3}$	$A_4 = 1.0448346 \cdot 10^{-4}$
$A_5 = 3.2673714 \cdot 10^{-2}$	$A_6 = -3.6453277 \cdot 10^{-3}$	$A_7 = 7.4299951 \cdot 10^{-5}$	$A_8 = -1.1381195 \cdot 10^{-1}$
$A_9 = 6.2476497 \cdot 10^{-4}$	$A_{10} = -1.0716866 \cdot 10^{-6}$	$A_{11} = 1.746622 \cdot 10^1$	

$$\ln P_d = A_1 [x_{C_2} + x_{CO_2} + x_{H_2S} + x_{C_6} + 2(x_{C_3} + x_{C_4}) + x_{C_5} + 0.4x_{C_1} + 0.2x_{N_2}] \quad (1)$$

$$+ A_2 \gamma_{C_{7+}} + A_3 \left(\frac{x_{C_1}}{X_{C_{7+}} + 0.002} \right) + A_4 T_R + A_5 (x_{C_{7+}} M_{C_{7+}})$$

$$+ A_6 (x_{C_{7+}} M_{C_{7+}})^2 + A_7 (x_{C_{7+}} M_{C_{7+}})^3 + A_8 \left(\frac{M_{C_{7+}}}{\gamma_{C_{7+}} + 0.0001} \right)$$

$$+ A_9 \left(\frac{M_{C_{7+}}}{\gamma_{C_{7+}} + 0.0001} \right)^2 + A_{10} \left(\frac{M_{C_{7+}}}{\gamma_{C_{7+}} + 0.0001} \right)^3 + A_{11}$$

Elsharkawy M. (2002)

$$P_d = A_0 + A_1 T_R + A_2 x_{H_2S} + A_3 x_{CO_2} + A_4 x_{N_2} + A_5 x_{C_1} + A_6 x_{C_2} + A_7 x_{C_3} + A_8 x_{C_4} \quad (2)$$

$$+ A_9 x_{C_5} + A_{10} x_{C_6} + A_{11} x_{C_{7+}} + A_{12} M_{C_{7+}} + A_{13} \gamma_{C_{7+}} + A_{14} (x_{C_{7+}} M_{C_{7+}})$$

$$+ A_{15} \left(\frac{M_{C_{7+}}}{\gamma_{C_{7+}}} \right) + A_{16} \left(\frac{x_{C_{7+}} M_{C_{7+}}}{\gamma_{C_{7+}}} \right) + A_{17} \left(\frac{x_{C_{7+}}}{x_{C_1} + x_{C_2}} \right)$$

$$+ A_{18} \left(\frac{x_{C_{7+}}}{x_{C_3} + x_{C_4} + x_{C_5} + x_{C_6}} \right)$$

Table A.2. The coefficients of Elsharkawy correlation.

A ₀ = 4268.85	A ₁ = 0.094056	A ₂ = -7157.87	A ₃ = -4540.58	A ₄ = -4663.55
A ₅ = -1357.56	A ₆ = -7776.10	A ₇ = -9967.99	A ₈ = -4257.10	A ₉ = -1417.10
A ₁₀ = 691.5298	A ₁₁ = 40660.36	A ₁₂ = 205.26	A ₁₃ = -7260.32	A ₁₄ = -352.413
A ₁₅ = -114.519	A ₁₆ = 8.133	A ₁₇ = 94.916	A ₁₈ = 238.252	

Shokir El-M. (2008)

$$P_d = B_1 + B_2 + B_3 + B_4 \quad (3)$$

$$B_1 = 201.875481x_{C7+} \left((T_R[x_{C3} - x_{H2S} + x_{CO2} - x_{C6} + x_{CO2} - x_{C4} - x_{C2}] \right. \\ \left. - [x_{C4}(x_{CO2} - x_{C4} - M_{C7+} + x_{N2} - x_{C5}M_{C7+}^2) - x_{C7+}] \right) \\ \left. - (x_{H2S} - x_{N2}T_R[x_{C1}^2 - x_{C7+}] + M_{C7+} - x_{C2} + x_{H2S}) \right) \\ + 38456.87953x_{C6} \quad (4)$$

$$B_2 = (0.000007 T_R x_{N2}) [(x_{CO2} - M_{C7+} - x_{C7+})(T_R - M_{C7+} - x_{CO2} - T_R) \\ - (x_{H2S} - T_R)(M_{C7+}^2 - x_{C3}M_{C7+})] + 225500.9399x_{C6} \quad (5)$$

$$B_3 = 120586.9719 x_{C1} [x_{H2S}(x_{C3}x_{H2S} - x_{C5} - x_{C7+}) \\ - ((x_{C7+} - x_{C1})(x_{C7+} - x_{C6}) - x_{H2S}x_{N2}^2)] + 72.6908M_{C7+} \quad (6)$$

$$B_4 = -1962.40851 x_{C5}(M_{C7+} - x_{C1}^2) \\ - 253385.67764x_{C7+}x_{CO2}x_{C3}(x_{CO2}x_{C3} - x_{C4} + x_{C7+})(x_{C3} - M_{C7+}) \\ - 13358.59271 x_{C4} + 4676.933602x_{C2} - 6567.9 \quad (7)$$

Ahmadi-Elsharkawy (2017)

$$P_d = -888.278 - (3.60639T_Rx_{C1}) + (0.00785623T_R^2) + (1467.87x_{C1}) \\ + (0.989073A) \quad (8)$$

$$A = 140.909 - \left(0.484983Bx_{N2}^{\frac{1}{3}}\right) + \left(0.576219Cx_{N2}^{\frac{1}{3}}\right) + \left(1746.2x_{N2}^{\frac{2}{3}}\right) \\ + (0.290811B) + (3.3387x10^{-5}B^2) + (0.484502C) \quad (9)$$

$$B = -6991.4 + (0.00165952M_{C7+}^3) - \left(0.00122815M_{C7+}^3x_{C7+}^{\frac{1}{3}}\right) \\ - (5.7182x10^{-11}M_{C7+}^6) - \left(79241x_{C4}^{\frac{1}{3}}x_{C7+}^{\frac{1}{3}}\right) + \left(31517.4x_{C4}^{\frac{2}{3}}\right) \\ + \left(50917.8x_{C7+}^{\frac{1}{3}}\right) - \left(31614.2x_{C7+}^{\frac{2}{3}}\right) \quad (10)$$

$$C = 29014 - (52127.9x_{C1}) + (79848.3\gamma_{C7+}^3) + \left(12633.6x_{C1}x_{C7+}^{\frac{1}{3}}\right) + (11116.5x_{C1}^2 - 58526.6\gamma_{C7+}^3) + \left(58263.6\gamma_{C7+}^3x_{C7+}^{\frac{1}{3}}\right) - \left(43792.7x_{C7+}^{\frac{2}{3}}\right) \quad (11)$$

Nnadozie O. (2018)

$$P_d = A_0 + A_1B_1 + A_2B_2 + A_3B_3 + A_4B_4 + A_5B_5 + A_6B_6 + A_7B_7 + A_8B_8 + A_9B_9 + A_{10}B_{10} + A_{11}B_{11} + A_{12}B_{12} + A_{13}B_{13} + A_{14}B_{14} + A_{15}B_{15} \quad (12)$$

Table A.3. The A_i coefficients of Nnadozie correlation.

$A_0 = 1.0$	$A_1 = 8.5$	$A_2 = 9.4$	$A_3 = 9.9$
$A_4 = 11$	$A_5 = 19.35$	$A_6 = 22.35$	$A_7 = 22.3$
$A_8 = 19.08$	$A_9 = 19.1$	$A_{10} = 19.1$	$A_{11} = 19.09$
$A_{12} = 19.1$	$A_{13} = 19.1$	$A_{14} = 19.12$	$A_{15} = 5$

Table A.4. The B_i coefficients of Nnadozie correlation.

$B_1 = T_R$	$B_2 = x_{H2S} + x_{CO2} + x_{N2}$	$B_3 = x_{C1}$	$B_4 = x_{C2}$
$B_5 = x_{C3}$	$B_6 = x_{C4}$	$B_7 = x_{C5}$	$B_8 = x_{C6}$
$B_9 = x_{C7+}$	$B_{10} = x_{C7+} M_{C7+}$	$B_{11} = M_{C7+}$	$B_{12} = \gamma_{C7+}$
$B_{13} = \frac{x_{C7+} M_{C7+}}{\gamma_{C7+}}$	$B_{14} = \frac{x_{C7+}}{x_{C1} + x_{C2}}$	$B_{15} = \frac{x_{C7+}}{x_{C3} + x_{C4} + x_{C5} + x_{C6}}$	

REFERENCES

- [1] Ahmed, T., Reservoir Engineering Handbook; Gulf Professional Publishing, USA. 2000.
- [2] Arabloo, M., Rafiee-Taghanaki, S., SVM modeling of the constant volume depletion (CVD) behavior of gas condensate reservoirs; Journal of natural gas science and engineering, vol. 21, pp. 1148-1155, 2014. <https://doi.org/10.1016/j.jngse.2014.11.002>.
- [3] Esmaeili-Jaghdan, Z., Tatar, A., Shokrollahi, A., Bon, J., Machine learning modelling of dew point pressure in gas condensate reservoirs: application of decision tree-based models; Neural Computing and Applications, vol.36, issue 05; pp. 1-23. (2024). <https://doi.org/10.1007/s00521-023-09201-9>.
- [4] Changjun, L., Yang, P., Jingya, D., Lei, C., Prediction of the dew point pressure for gas condensate using a modified Peng-Robinson equation of state and Four-coefficient molar distribution function; Journal of Natural Gas Science and Engineering, vol. 27 Part 2, pp. 967-978. 2015. <https://doi.org/10.1016/j.jngse.2015.09.030>.
- [5] McCain, W.D., The properties of petroleum fluids; PennWell Publishing Company, USA. 1990.

- [6] Bennion, D.B., Thomas, F.B., Schulmeister, B., Retrograde Condensate Dropout Phenomena in Rich Gas Reservoirs-Impact on Recoverable reserves, Permeability, diagnosis and simulation techniques; *Journal of Canadian Petroleum Technology*, vol. 40, issue 12, Calgary. 2001. <https://doi.org/10.2118/01-12-tn1>.
- [7] Nnadozie, O., A new analytical method for predicting dew point pressure for gas condensate reservoirs; SPE-162985, Nigeria. 2012. <https://doi.org/10.2118/162985-MS>.
- [8] Potsch, K.T., Braeuer, L., A novel graphical Method for determining Dewpoint pressure of gas Condensates; SPE 36919, Milan. 1996. <https://doi.org/10.2118/36919-MS>.
- [9] Adetiloye, B., Determination of the Dew-Point Pressure (DPP) for a gas Condensate Fluid by genetic Algorithm (GA); SPE-167620-STU, New Orleans. 2013.
- [10] Zeinab A., Chis T., Pozo A.O., A Short Review about Predicting the Dew Point Pressure for Gas Condensate Reservoirs by Empirical Correlations, *International Journal of Engineering Research and Reviews*, Vol. 11, Issue 3, pp. 31-40, July-September 2023.
- [11] Zeinab A., Pozo A.O., Brezeanu M.C., Chis T., Novel acid soluble abrasive material for abrasive jet perforation through coiled tubing and TS-tool for carbonate gas well – Yard test, *Romanian Journal of Petroleum & Gas Technology*, 4 (2), 2023, pp. 97-108, <https://doi.org/10.51865/JPGT.2023.02.10>.
- [12] Elsharkawy, A.M., Predicting the dew point pressure for gas condensate reservoirs: empirical models and equations of state; *Fluid phase equilibria*, vol. 4900; pp. 1-20, Kuwait. 2001. [https://doi.org/10.1016/S0378-3812\(01\)00755-6](https://doi.org/10.1016/S0378-3812(01)00755-6).
- [13] Shokir, E., Dewpoint pressure model for gas condensate reservoirs based on genetic programing; SPE-114454, Alberta. 2008. <https://doi.org/10.2118/114454-MS>.
- [14] Nemeth, L.K., Kennedy, H.T., A Correlation of Dewpoint Pressure with Fluid Composition and Temperature; SPE-1477-PA. 1967. <https://doi.org/10.2118/1477-PA>.
- [15] Shokrollahi, A., Arabloo, M., Evolving an accurate model based on machine learning approach for prediction of dew-point pressure in gas condensate reservoirs; *chemical engineering research and design*, vol. 92, pp. 891-902, Iran. 2013. <https://doi.org/10.1016/j.cherd.2013.08.019>.
- [16] Ahmadi, M.A., Elsharkawy, A., Robust correlation to predict dew point pressure of gas condensate reservoirs; *Petroleum*, vol. 30, pp. 1-8. 2016. <https://doi.org/10.1016/j.petlm.2015.10.005>.
- [17] Rabiei, A., Sayyad, H., Riazi, M., Determination of dew point pressure in gas condensate reservoirs based on a hybrid neural genetic algorithm; *Fluid Phase Equilibria*, vol. 387, pp. 38-49. 2014. <https://doi.org/10.1016/j.fluid.2014.11.011>.
- [18] Arabloo, M., Shokrollahi, A., Gharagheizi, F., Mohammadi, A.H. (). Toward a predictive model for estimating dew point pressure in gas condensate systems; *Fuel Processing Technology*, vol. 116, pp. 317-324. 2013. <https://doi.org/10.1016/j.fuproc.2013.06.012>.
- [19] El-Aily, M., Khalil, M.H.M., Desouky, S.M., Batanoni, M.H., Mahmoud, M.R.M., Experimental studies on constant mass–volume depletion of gas-condensate systems; *Egyptian Journal of Petroleum*, vol. 22, issue 1, pp. 129-136. 2013. <https://doi.org/10.1016/j.ejpe.2012.09.006>.
- [20] Rojas, G., *Ingeniería de Yacimientos de Gas Condensado*; Ediciones del Departamento del Petroleo de la Universidad de Oriente, Venezuela. 2023.

-
- [21] Fan, L., Harris, B.W., Understanding gas condensate reservoirs; Chevron Oil Field Review. 2005.
 - [22] Hussam, M., Shokir, E., Sayyauh, M.H., Prediction of the PVT data using Neural Network Computing Theory; SPE-85650, Nigeria. 2003. <https://doi.org/10.2118/85650-MS>.
 - [23] Nasrifar K., Bolland, O., Predicting Natural Gas Dew Points from 15 Equations of State; Energy and Fuels; vol. 19, pp. 561-572. 2004. <https://doi.org/10.1021/ef049806s>.
 - [24] Kaydani, H., Mohebbi, A., Hajizadeh, A., Dew point pressure model for gas condensate reservoirs based on multi-gene genetic programming approach; Applied Soft Computing, vol. 47, pp. 168-178, Iran. 2016. <https://doi.org/10.1016/j.asoc.2016.05.049>.
 - [25] Lertliangchai, T., Dindoruk, B., Lu, L., Yang, X., A Comparative Analysis of the Prediction of Gas Condensate Dew Point Pressure Using Advanced Machine Learning Algorithms; Fuels, vol. 5, issue 3, pp. 548-563. 2024. <https://doi.org/10.3390/fuels5030032>.

Received: May 2026; Revised: June 2026; Accepted: June 2026; Published: June 2026