International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 2 February - 2026, Pages: 78-84

Predictive Modeling of Glycol Recovery Efficiency in Natural Gas Processing Using Machine Learning

Christian Emelu Okalla[1]*, Charley Iyke C. Anyadiegwu[1], Ubanozie Julian Obibuike[1], Prospere Chimdike Nwokorie[1], Chukwudi Michael Ohaegbulam[1], & Chika Veronica Nwachukwu[1]

Department of Petroleum Engineering
Federal University of Technology Owerri
Imo State, Nigeria
*christian.okalla@futo.edu.ng

ABSTRACT— *Gas dehydration is an important aspect of gas processing since it tends to remove water from natural gas there-by minimizing the formation of gas hydrates during transportation in pipelines. Gas dehydration is usually conducted using triethylene glycol (TEG) which is usually recovered after the process. It is important to note that the recovered TEG should have a high purity so that when reused, it can be more effective. A solution to this problem is proposed in this study such that a model is developed which can be used to predict TEG purity and hence define specific input values to the system to achieve the desired purity. The model was developed using Design Expert Software with input data from design of experiments and output data from Aspen Hysys simulation of TEG purity. Analysis of Variance was conducted to determine the model significance and model terms that are significant to the model. The developed model was validated using cross plots and trend analysis which proved to be valid. This is because the data points were located on the 45° line and also the actual and predicted results followed the same pattern depicted by overlap of the curves. The model developed in this study can be used to predict TEG purity for given values of TEG flow rate, column pressure, and reboiler temperature. This work can aid in reducing costs associated with performing trial and error simulations with the real system or with the simulation software.*

Keywords: Gas dehydration, Glycol, TEG, Dehydration, Natural Gas Processing

1. INTRODUCTION

Natural gas is one of the world's most important sources of clean energy due to its relatively low carbon emissions, high energy content, and wide range of industrial applications [1, 2]. However, natural gas as produced from the reservoir is rarely pure; it typically contains impurities such as water vapor, carbon dioxide (CO_2), hydrogen sulfide (H_2S), and heavier hydrocarbons [3, 4]. Among these impurities, water vapor poses a significant challenge as it can lead to the formation of gas hydrates and corrosion in pipelines and processing equipment [5, 6]. To prevent these operational issues, gas dehydration is a crucial step in natural gas processing, and one of the most effective methods employed is glycol dehydration [7, 8]. Glycol dehydration involves the use of hygroscopic liquid desiccants, primarily Triethylene Glycol (TEG), Diethylene Glycol (DEG), or Monoethylene Glycol (MEG), to absorb water from wet natural gas streams [9, 10]. During the process, glycol circulates between the absorber and the regenerator, allowing for continuous removal and recovery of water vapor [11]. The efficiency of glycol recovery determines both the economic performance and environmental sustainability of the system [12, 13]. High glycol recovery minimizes operational costs, reduces losses, and enhances process safety [14]. However, achieving optimal recovery efficiency depends on several factors such as operating temperature, pressure, glycol circulation rate, lean glycol purity, and column design parameters [15, 16].

Traditionally, the modeling and optimization of glycol recovery efficiency have relied on empirical correlations and process simulation software such as Aspen HYSYS or PRO/II [17, 10]. While these approaches provide useful insights, they are often limited by their reliance on idealized assumptions, high computational costs, and the inability to accurately represent nonlinear relationships inherent in real-world operations [8, 6]. In recent years, Machine Learning (ML) techniques have emerged as powerful tools for process modeling and prediction in the oil and gas industry [11]. ML algorithms can learn complex patterns from historical process data, enabling the development of predictive models that are both robust and adaptable to varying operational conditions [12]. Despite the importance of glycol recovery in ensuring process efficiency and cost-effectiveness, many gas processing plants still face challenges in achieving optimal recovery performance. Traditional models often fail to capture the nonlinear interactions between process parameters, leading to suboptimal predictions and operational inefficiencies. Furthermore, the lack of data-driven predictive frameworks limits proactive decision-making in glycol regeneration and process control. There is, therefore, a need to develop a machine learning-based predictive model that can accurately estimate glycol recovery efficiency and assist operators in optimizing process parameters in real time. Natural gas is a crucial global energy source composed primarily of hydrocarbon gases like methane. While valued for its cleanliness and safety, raw natural gas extracted from reservoirs contains impurities that make it unsuitable for direct use. These impurities include water vapor and acid gases like carbon dioxide (CO_2) and hydrogen sulfide (H_2S), which can cause significant operational problems such as pipeline corrosion and the formation of solid hydrates under high-pressure, low-temperature conditions [1, 6]. Therefore, processing is an essential intermediary step between extraction and consumer delivery to meet quality and safety standards [17].

The processing chain involves several key purification stages. Initially, raw gas is separated from liquids like oil and water at the production site. It then undergoes sweetening to remove corrosive acid gases using chemical solvents like amines. A subsequent critical step is dehydration, which eliminates water vapor using liquid desiccants (e.g., triethylene glycol) or solid desiccants (e.g., molecular sieves) to prevent hydrate formation and blockages [2, 8]. Finally, the gas enters fractionation, where heavier hydrocarbon components (Natural Gas Liquids) such as ethane and propane are recovered as valuable byproducts. The resulting "pipeline-quality gas" is predominantly methane (85–95%) [3, 16].

The effectiveness of these processes depends on careful design, operational parameters, and control strategies. Modern plants utilize advanced modeling and simulation for optimization, and there is growing interest in employing machine learning and data-driven predictive models to enhance real-time performance and efficiency [8, 7]. Polák [10] applied mechanistic modeling using process simulation software (NTNU) to model the absorption drying process for natural gas. The key contribution of this work was the demonstration that such simulators can effectively replicate real plant behavior. Additionally, the research provided an important critique by identifying specific limitations inherent in conventional equilibrium-based modeling approaches for this application. Ahmad [8] investigated natural gas dehydration via Triethylene Glycol (TEG) through a combined experimental and simulation approach. The study's principal contribution was the identification and analysis of critical operational parameters specifically, temperature, pressure, and glycol circulation rate that directly govern the efficiency of glycol recovery in the dehydration process. Tk et al. [12] focused on the glycol regeneration unit to examine the effect of stripping gas on Triethylene Glycol (TEG) losses. The key finding of their work showed that strategically adjusting the stripping gas flow rate is a critical control measure, as it directly improves the efficiency of glycol recovery while simultaneously decreasing loss rates in the dehydration process. Gandhidasan [9] performed a detailed parameter analysis of Triethylene Glycol (TEG) dehydration systems using theoretical and computational modeling. The principal contribution of this research was the development of predictive correlations that quantify the relationship between key operational parameters and the resulting dehydration efficiency. The above existing researches has firmly established the foundational principles involving mechanistic and equilibrium-based models and theoretical correlations which explain system behavior, while simulation and experimental studies have isolated key influential parameters (temperature, pressure and circulation rate). However, these approaches possess inherent limitations. They often rely on steady state assumptions and may not fully capture the complex, interdependent effects of multiple variables changing simultaneously in an operational plant. Machine learning presents a paradigm shifting solution to this gap. A machine learning based predictive model can consume real time historical plant data encompassing all parameters previously studied in isolation. This would enable the prediction of glycol recovery efficiency under varying conditions.

2. Materials and Methods

2.1 Materials

The materials used in this study include Aspen Hysys and Design Expert Software.

1. Aspen Hysys was used to perform TEG recovery simulations for different operating conditions including TEG flow rate, column pressure, and reboiler temperature to get TEG purity.
2. Design Expert Software was used in modeling the simulation data to generate a relationship between TEG recovery purity referred to as the dependent variable and TEG flow rate, column pressure, and reboiler temperature referred to as the independent variables.

2.2 Method

Design of experiments feature was used to generate parameter realizations for conducting TEG recovery simulations. The Box Behnken experimental design method was used to generate 12 input parameter realizations. The input parameter realizations were used as input to Aspen Hysys to perform TEG recovery simulations. The output from simulations is referred to as TEG purity which is expressed as a percentage.

The input data obtained from Design Expert and the TEG purity results from Aspen Hysys for each input parameter realization are entered and further for analyzed with Design Expert to aid in model development. In the first instance, the analyst selects a transformation from a drop down as shown in figure 1. Usually at the onset, a no transformation is used such that later within the analysis, a recommendation is made on the most suitable transformation function for the data. When a transformation function is selected, analysis is started which is followed by fit summary. In this section, the software suggests a suitable model type for the data used in the study. The fit summary is shown in Figure 2.

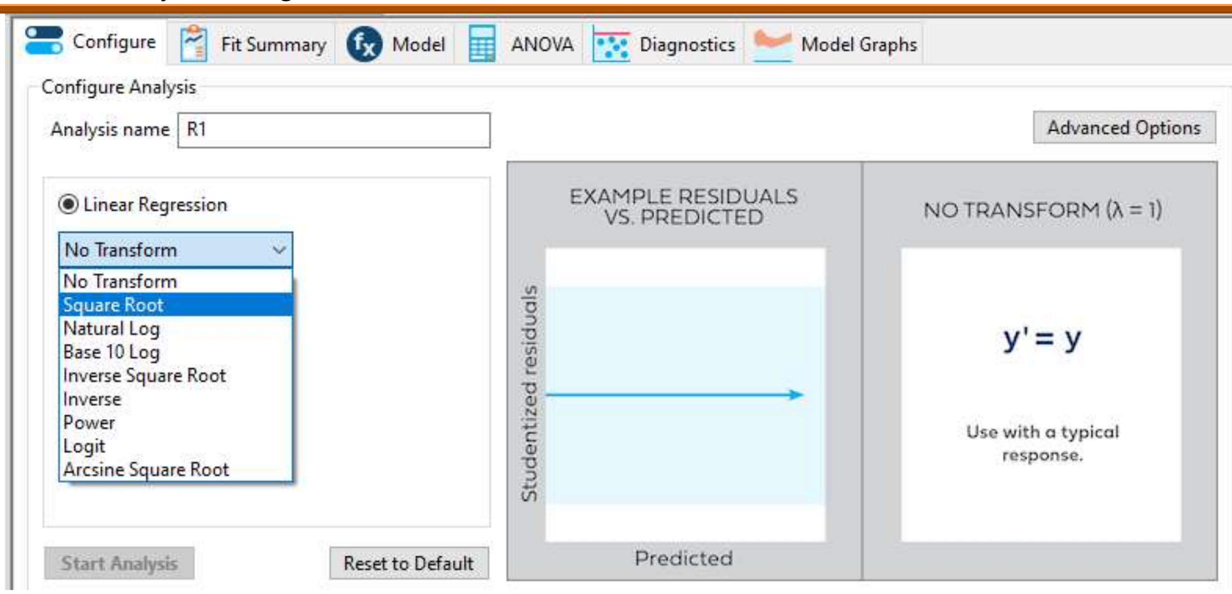

Figure 1: Configuration of Model

The fit summary tab is clicked to determine the model type. This is shown in Figure 2.

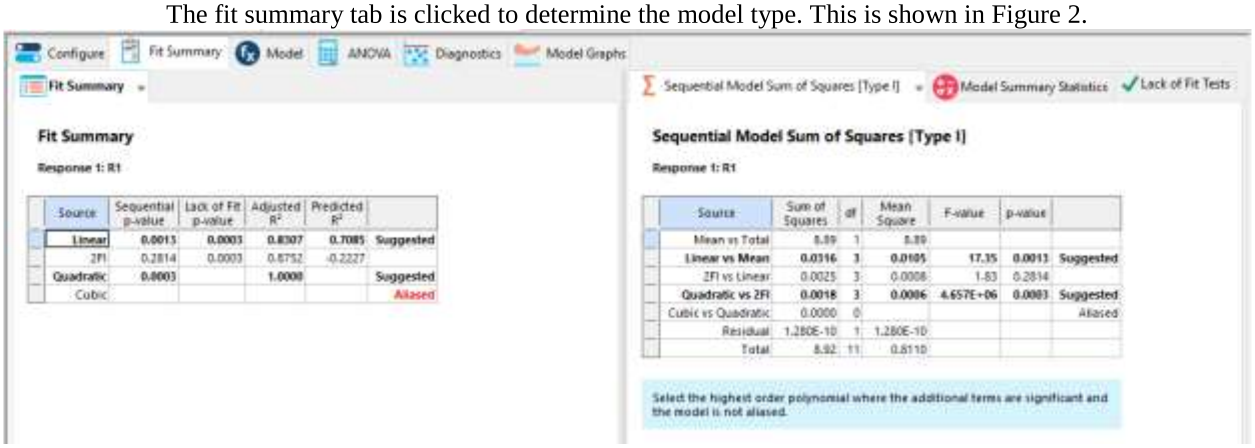

Figure 2: Fit Summary Tab

From the fit summary, the most suitable model type is highlighted, from which the modeling is conducted, followed by Analysis of Variance (ANOVA).

In this section, the model and ANOVA tabs are interchanged till a model with suitable accuracy is obtained. The developed model is validated using cross plots and trend analysis. It is expected for the developed model to be considered as being valid or accurate if the data points fall around the 45° line and if the actual and predicted results follow the same pattern.

3. Results and Discussion

Design of experiments was conducted using Box Behnken design method to generate 11 parameter realizations which were used in Aspen Hysys to perform series of simulations for TEG purity. The input and output data from Design of experiments and Aspen Hysys is shown in Table 1.

Table 1: Input and Output data from design of experiment and Aspen Hysys

Index	TEG Flow Rate (Kgmol/h)	Column Pressure (psi)	Reboiler Temperature (°F)	TEG Purity (fraction)
Symbol	A	B	C	R1
1	2.8	40	136	0.783822
2	2.8925	55	157.5	0.898338
3	2.985	63.2	175	0.905992

4	3.0125	76.25	196.25	0.946879
5	3.1	85.9	205	0.96488
6	2.91	40.52	190.97	0.965133
7	3.1	85.9	191	0.921213
8	2.87	40.5	152.71	0.904416
9	2.87	40.5	152.71	0.904432
10	2.87	40.5	150	0.896881
11	2.97	85.9	150	0.794419

The data shown in Table 1 was further analyzed with design expert software. A linear and quadratic model types were suggested based on the calculated R2, predicted R2, and adjusted R2 values as shown in Table 2.

Table 2: Fit Summary for TEG purity model

Source	Sequential p-value	Lack of Fit p-value	Adjusted R^2	Predicted R^2	
Linear	**0.0013**	**0.0003**	**0.8307**	**0.7085**	**Suggested**
2FI	0.2814	0.0003	0.8752	-0.2227	
Quadratic	**0.0003**		**1.0000**		**Suggested**
Cubic					**Aliased**

Based on the fit summary, the linear type of model was selected and Analysis of Variance (ANOVA) was conducted and the results are presented in Table 3.

Table 3: Analysis of Variance for the Quadratic TEG Purity Model

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	0.0358	9	0.0040	3.109E+07	0.0001	significant
A-A	0.0003	1	0.0003	2.091E+06	0.0004	
B-B	0.0003	1	0.0003	2.256E+06	0.0004	
C-C	0.0001	1	0.0001	6.075E+05	0.0008	
AB	0.0007	1	0.0007	5.160E+06	0.0003	
AC	0.0013	1	0.0013	1.005E+07	0.0002	
BC	0.0008	1	0.0008	6.371E+06	0.0003	
A^2	0.0008	1	0.0008	6.404E+06	0.0003	
B^2	0.0005	1	0.0005	4.119E+06	0.0003	
C^2	9.352E-07	1	9.352E-07	7306.60	0.0074	
Pure Error	1.280E-10	1	1.280E-10			
Cor Total	0.0358	10				

The ANOVA results show that the overall model and lack of fit are significant. Hence, the model for predicting TEG purity can be outputted. The model is presented in Equation 1

$$\boldsymbol{TEG\ Purity = -777.632 + 605.996 * A + -2.44709 * B + -0.477557 * C + 0.958444 * AB + 0.17885 * AC + -0.000930861 * BC + -117.774 * A\hat{}2 + -0.00179206 * B\hat{}2 + 1.56649e-05 * C\hat{}2} \quad (1)$$

Where A = TEG Flow Rate, B = Column Pressure, and C = Reboiler Temperature

The model presented by equation 1 was validated using cross plots and trend analysis. The cross plot and trend analysis results are shown in Figures 3 and 4 respectively.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 2 February - 2026, Pages: 78-84

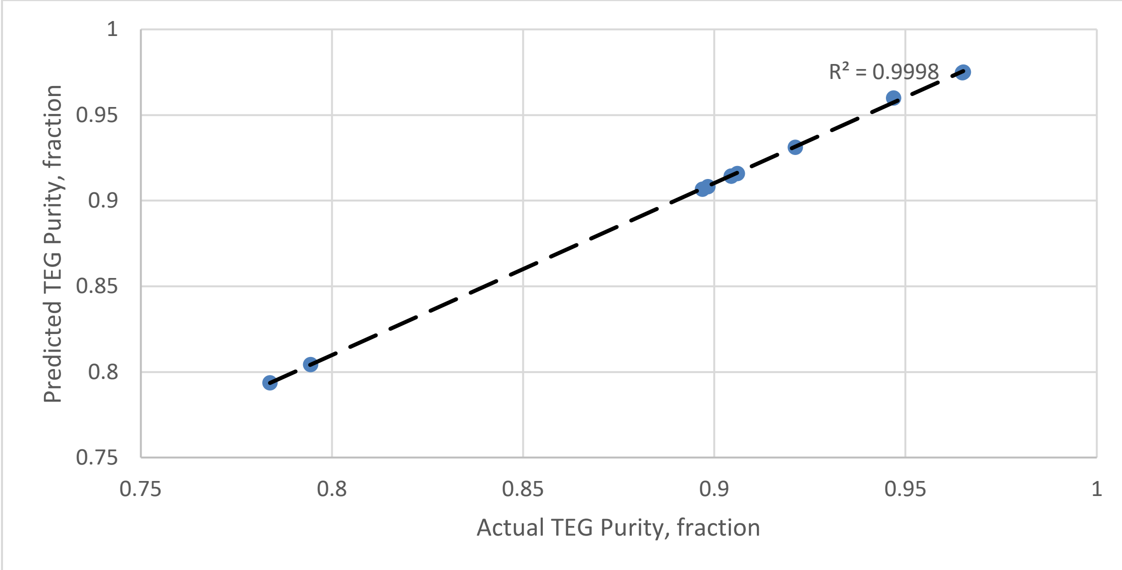

Figure 3: Cross Plots for TEG Purity

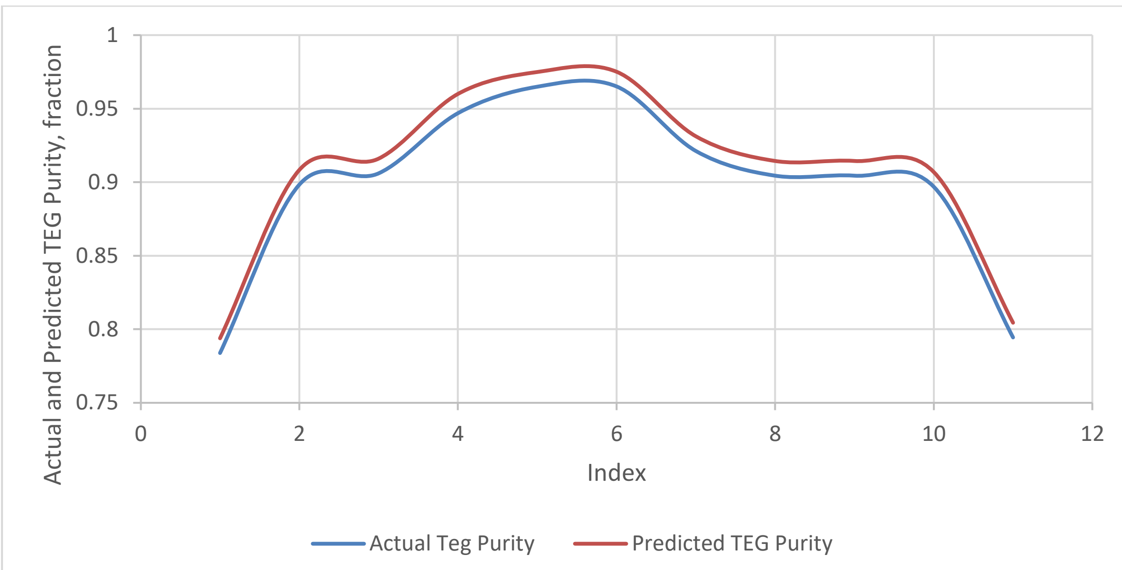

Figure 3: Trend analysis for TEG Purity

In this study, Aspen Hysys was used to simulate triethylene glycol (TEG) recovery after gas dehydration using data generated with design of experiments found in Design Expert software. With the aid of DE software, 11 parameter realizations were generated for three input parameters namely triethylene glycol (TEG) flow rate, column pressure, and reboiler temperature. The generated parameter realizations were used as input to Hysys for TEG recovery and 11 corresponding TEG purity results were obtained. The data shown in Table 1 were entered into Design Expert software for further analysis and model development. Fit summary was conducted which determined the most suitable model type for the data used in this study (Table 2). From the fit summary, the quadratic model type was selected. Analysis of Variance (ANOVA) was conducted on the selected model type and the results are presented in Table 3. ANOVA results show that the overall model and the model terms are significant since their p-values are less than 0.05. Since the model is significant, the TEG purity model was outputted as shown in Equation 1. The model was validated using cross plots and trend analysis which depicts validity. This is because the plotted points appeared clustered or were found on the 45° line (Figure 3) while the actual and predicted TEG purity results were found to follow the same pattern depicted by an overlap of the actual and predicted TEG purity curves (Figure 4).

4. Conclusion

In this study, the following conclusions were arrived at;

a. A model for predicting TEG purity was developed as shown in Equation 1. The independent variables to the model include TEG flow rate, Column pressure, and reboiler temperature.
b. The developed model can be helpful in predicting TEG purity since the input parameter values can easily be tuned to get a desired purity.
c. The model developed in this study is limited only to the data range used for its development.

5. Recommendation

The following recommendations have been proposed based on the results and insights from this study

a. More data points need to be obtained from simulations and also from industrial processes to enable an increase in data range and wider operational capability
b. Other types of machine learning algorithms such as artificial neural networks, Gaussian process regression, and support vector regression can be used to develop the models as well but will require massive data to achieve this.

Conflicting Interests

The author(s) declare that they have no conflicting interests.

6. References

[1] Lyons, W. C. (1994). Standard Handbook of Petroleum and Natural Gas Engineering. Vol. 3.

[2] Katz, D. L. (1990). Handbook of Natural Gas Engineering, Production and Storage. McGraw-Hill Book Co., Inc.

[3] Arnold, K. & Stewart, M. (1999). "Surface Production Operations", Vol 2, Design of Gas Handling Systems and Facilities.

[4] Guo, B. & Ghalambor, A. (2005). Natural Gas Engineering Handbook. Gulf Publishing Company, Houston Texas, USA.

[5] Ikoku, C. U. (1992). Natural Gas Production Engineering. Reprint Edition, Krieger Publishing Company: Boca Raton, FL.

[6] GPSA. (2004). Engineering Data Book. 12th ed. Gas Processors Suppliers Association (GPSA). Tulsa: GPSA Press.

[7] Etuk, P. (2007). Total E&P Gas Dehydration Training Manual course EXP-PR-PR130. Rev.01999.1.

[8] Ahmad. (2009). "Natural Gas Dehydration using Triethylene Glycol (TEG)", Publication of the University of Malaysia Pahang, April.

[9] Gandhidasan, P. (2003). Parameter Analysis of Natural Gas Dehydration by Triethylene Glycol Solution. Energy 26: 855-201.

[10] Polák, L. (2009). Modeling Absorption drying of natural gas", Norwegian University of Science and Technology (Norwegian: Norges Teknisk-Naturvitenskapelige University (NTNU)) Trondheim, Norway.

[11] Guo, B., Lyons, W. C., & Ghalambor, A. (2007). Petroleum Production Engineering: A computer Assisted Approach. Elsevier Science & Technology Books.

[12] Tk, I., Rk, A., Fh, K., & Im, K. (2017). The Impact of Stripping Gas Flow Rate on Triethylene Glycol Losses from Glycol Regeneration Unit: Simulation Study. Journal of Chemical Engineering & Process Technology, 08(03). https://doi.org/10.4172/2157-7048.1000337

[13] Pimchanok. (2013). Membrane Based Triethylene Glycol Separation and Recovery from Gas Separation Plant Wastewater. Asian School of Technology, Thailand.

[14] Abdel-Aal, H. K., Aggour, M. A., & Fahim, M. A. (2003). Petroleum and Gas Field Processing. Marcel Dekker, Inc.

[15] Gandhidasan, P, Al-Farayedhi, A., & Al-Mubarak, A. (2001). Dehydration of natural gas using solid desiccants. Energy 26: 855-868.

[16] Mokhatab, S. (2010). An optimum lineup for sour gas processing. Petrol. Technol. Quart. 37--43. Gas Issue

[17] Mokhatab, S., & Meyer, P. (2009). Selecting best technology lineup for designing gas processing units. Paper Presented at Gas Processors Association (GPA) Europe Sour Gas Processing Conference, Sitges, Barcelona, Spain.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 2 February - 2026, Pages: 78-84

Authors

Christian Emelu Okalla, is a Technologist and Researcher at the Department of Petroleum Engineering, Federal University of Technology Owerri, Nigeria. He holds both B.Eng and M.Eng in Petroleum Engineering from the Federal University of Technology Owerri. His research interests are in drilling engineering, production engineering, natural gas engineering, reservoir engineering, and reservoir simulation.

Charlie Iyke Anyadiegwu is an Associate Professor in the Department of Petroleum Engineering, Federal University of Technology Owerri where he has worked for over 25 years. He holds both B.Eng and M.Eng in Petroleum Engineering and Gas Engineering Respectively from the University of Portharcourt, Rivers State, Nigeria, and Ph.D. degree in Petroleum Engineering from the Federal University of Technology Owerri, Imo State, Nigeria. His research interests are in oil and gas production and processing from fossil fuel and non-fossil fuel (biomass), health, safety and environment (HSE), oil spillage detection, control and prevention.

Ubanozie Julian Obibuike is a Senior Lecturer with over 17 years of invaluable teaching and research experience at the Department of Petroleum Engineering, Federal University of Technology Owerri. Dr. Obibuike's research interest covers a wide array of pivotal areas such as pipeline performance integrity, advance reservoir simulation and development techniques, petroleum economic analysis, fluid flow dynamics in natural gas systems and innovative drilling technologies. Dr. Obibuike holds both PhD and M.Eng in Petroleum Engineering from the Federal University of Technology Owerri..

Prospere Chimdike Nwokorie is a recent graduate of the Department of Petroleum Engineering, Federal University of Technology Owerri. He holds a B.Eng in Petroleum Engineering from the Federal University of Technology Owerri, Imo State, Nigeria. His research interests are in drilling, production, and reservoir Engineering.

Chukwudi Michael Ohaegbulam is a petroleum engineer and a lecturer at Madonna University Nigeria. He has over ten years of teaching and industrial experience. His expertise is drilling optimization, transport in porous media, reservoir modeling and simulation, multiphase flow in pipes and geothermal wells

Chika Veronica Nwachukwu is a COREN-certified Petroleum Engineer and Assistant Chief Technologist in the Department of Petroleum Engineering, Federal University of Technology Owerri (FUTO). An alumnus of the department, she holds a Master's degree in Petroleum Engineering and is currently pursuing an additional Master's degree in Environmental Standards.