Linear Regression And Piecewise Multiple Linear Regression Function Models For Prediction Of Crude Oil Storage Tank Volume

Emmanuel Udama Odeh[1]
Department of Mechanical and Aerospace Engineering,
University of Uyo, Uyo Akwa Ibom State Nigeria
emmanuelodeh@uniuyo.edu.ng

John Dennis Urua[2]
Department of Mechanical and Aerospace Engineering,
University of Uyo, Uyo Akwa Ibom State Nigeria
johnurua1@gmail.com

Aaron Tukpakonbowei Ben-Elijah [3]
Ralph E. Martin Department of Chemical Engineering,
University of Arkansas, USA
aaron.benelijah@gmail.com
ORCID No: 0009-0009-0245-1650

***Abstract*—This study compares Linear Regression (LR) and Piecewise Multiple Linear Regression Function (PMLRF) models for predicting crude oil storage tank volume based on Electro-Optical Distance Ranging (EODR) and Manual Strapping Method (MSM) data. While both methods achieved 99.99% accuracy, the PMLRF model outperformed the LR model, reducing maximum errors from 13.8777 to 2.5632 and yielding lower overall error metrics (MSE of 0.18770 vs 0.1997). Specifically, while both models achieved an identical prediction accuracy of 99.99% (expressed as 1-MAPE), the PMLRF model demonstrated much higher stability across different tank depths. The LR model exhibited a massive peak error of 13.8777 at a depth of 8210.0. In contrast, the PMLRF model's maximum error was only 2.5632 (at depth 2420), representing an 81.5% reduction in worst-case prediction error. Also, the LR model's minimum error was 0.0172 (at depth 1880.0), while the PMLRF model achieved a tighter minimum error of -0.0948 (at depth 2100). In all, the statistical improvement in the PMLRF model suggests it better captures the non-linear or segmented nature of crude oil storage tank volumes. By using a piecewise approach, the model adapts to different sections of the tank depth rather than applying a single global linear trend, which effectively minimized the large residual spikes seen in the standard LR model. Finally, the findings indicate the PMLRF model better accounts for non-linearities in tank calibration. For more details, consult the study's findings on the methods' predictive performance.**

Keywords—Linear Regression (LR) Model, Piecewise Multiple Linear Regression Function (PMLRF) Model, Ground Truthing, Electro-Optical Distance Ranging (EODR), Crude Oil Storage Tank Calibration Dataset

1. Introduction

The calibration of crude oil storage tanks is a critical process, mandated to ensure precise volume measurement for custody transfer [1,2]. The Conventional Manual Strapping Method (MSM) involves physical measurement of tank shells, a process that is time-consuming and labor-intensive. Furthermore, the inherent safety hazards associated with manual techniques, such as working at heights and in confined areas, necessitate safer, more modern alternatives [3].

The introduction of EODR represents a significant advancement in non-contact measurement, addressing the limitations of manual processes [4,5]. By eliminating the need for scaffolding and providing ground-based measurements, it enhances safety while delivering higher accuracy in shorter timeframes [6]. However, since historical records remain a cornerstone of the industry, a bridge between MSM and EODR is required. Establishing a predictive model to convert EODR data into MSM-equivalent values is vital to ensure that this technological transition does not compromise data continuity or accuracy [7].

Previous studies have shown that while EODR is efficient, converting its data into the precise format of the Manual Strapping Method (MSM) is required for standardized reporting [8,9]. Machine learning and regression techniques have been increasingly utilized to model complex relationships in oil production and storage, providing higher predictive performance than manual calculations [10,11,12]. This research aims to bridge the gap between traditional manual methods and modern technology by developing a predictive framework that translates EODR data to MSM equivalents. By utilizing Linear Regression (LR) and Piecewise Multiple Linear Regression Function (PMLRF) models on pairs of EODR and MSM data, this study seeks to create a robust, accurate, and efficient automated system for predicting crude oil tank volume, thereby eliminating the tedium of manual calibration.

2. Methodology

The primary objective of this methodology is to develop a predictive framework that transforms Electro-Optical Distance Ranging (EODR) data into Manual Strapping Method (MSM) equivalent values, effectively eliminating the need for tedious manual calibration. The study utilizes two distinct datasets derived from crude oil storage tank calibrations. The first is the Ground Truth Dataset (MSM) which is obtained through traditional manual strapping. While highly accurate, it serves as the benchmark for "true" volume. The second is the input dataset (EODR) which is obtained via electronic sensors. This serves as the independent variable (x) due to its ease of collection despite its inherent variance from the MSM values. The datasets are paired by tank depth, ensuring that each measured EODR volume corresponds to a specific MSM volume at the same vertical level.

Specifically, the methodology for this research involves a comparative analysis and predictive modeling approach to translate crude oil storage tank calibration data obtained from the Electro Optical Distance Ranging (EODR) method into the highly accurate format of the Manual Strapping Method (MSM). The goal is to develop machine learning models that accept EODR volume measurements and output the equivalent MSM volume for any given tank depth, thereby eliminating the need for tedious manual calibration.

Two main regression approaches are employed to establish the relationship between V_{EODR} and V_{MSM}. The first is the Linear Regression Model (LR) and the second approach is the Piecewise Multiple Linear Regression Function (PMLRF) Model.

2.1 Development of the Linear Regression Model

At its core, linear regression is used to map the connection between a target variable (Y) and its predictors (X). By calculating a best-fitting line that narrows the gap between reality and prediction, we can better understand the dynamics between depth measurements and fluid volume in the MSM/EODR datasets. This approach not only clarifies the impact of depth on volume but also serves as a tool for interpolating missing information. For a single predictor variable (e.g., depth vs volume increment), the equation is given by:

$$V = \beta_0 + \beta_1 \times D + \epsilon \qquad (1)$$

Where, V denotes fluid volume increment, D is the depth measured in mm, β_0 is the intercept which represent the predicted fluid volume when $D = 0$, β_1 is the slope, and ϵ is the residual error. For multiple predictors, such as using both MSM and EODR increments;

$$V = \beta_0 + \beta_1 \times D + \beta_2 \times I_{MSM} + \beta_3 \times I_{EODR} + \epsilon \qquad (2)$$

Where, I_{MSM} is the increment in MSM volume, and I_{EODR} is the increment in EODR volume.

Least squares method is used to compute β_0 and β_1, in order to minimize sum of squared residuals;

$$min\left(\sum_{i=1}^{n}\left(V_i - (\beta_0 + \beta_1 \times D_i)\right)^2\right) \tag{3}$$

Solving for β_1:

$$\beta_1 = \frac{\sum(D_i - \bar{D})(V_i - \bar{V})}{\sum(D_i - \bar{D})^2} \tag{4}$$

Then intercept can be computed as:

$$\beta_0 = \bar{V} - \beta_1 \cdot \bar{D} \tag{5}$$

These formulas allow effective computation of the regression coefficients from the dataset. Training the Model Over Epochs: Unlike ordinary least squares (OLS), when training using gradient descent, the parameters are updated iteratively over multiple epochs as:

$$\beta_j^{(t+1)} = \beta_j^t - \alpha \frac{\partial J}{\partial \beta_j} \tag{6}$$

Where, β_j^t is the parameter at epoch t, α is the learning rate, and J is the cost function (typically the mean squared error):

$$J(\beta) = \frac{1}{n}\sum_{i=1}^{n}(V_i - \bar{V}_i)^2 \tag{7}$$

As training progresses, MSE should decrease until convergence.

2.2 The Development of the Piecewise multiple linear regression function (PMLRF) model

The graph of the MSM dataset and the EODR dataset versus tank depth is shown in Figure 1 while the MSM dataset and the EODR dataset is shown in Figure 2. The two graphs showed that there is a linear relationship between the tank depth and the MSM dataset and also between the tank depth and the EODR dataset. Also, the two graphs showed that there is a linear relationship between tank depth the MSM dataset and the EODR dataset. In addition, there is a sharp difference in the gradient of the lines at the tank depth of 3050 mm in the two graphs.

Furthermore, the two trend line equations in Figure 2 were used to predict the values of the MSM data from the corresponding EODR data and the prediction error was plotted against the tank depth and the result is shown in Figure 3.

the prediction error graph is shown in Figure 3.3.

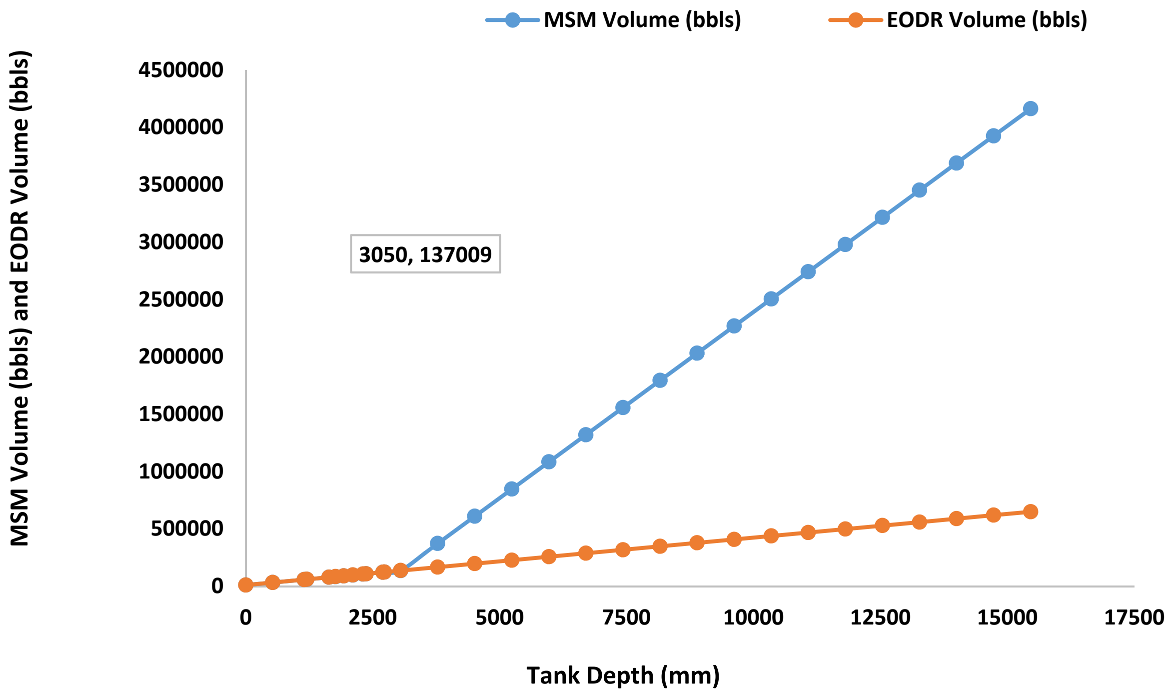

Figure 1 The graph of the MSM dataset and the EODR dataset versus tank depth

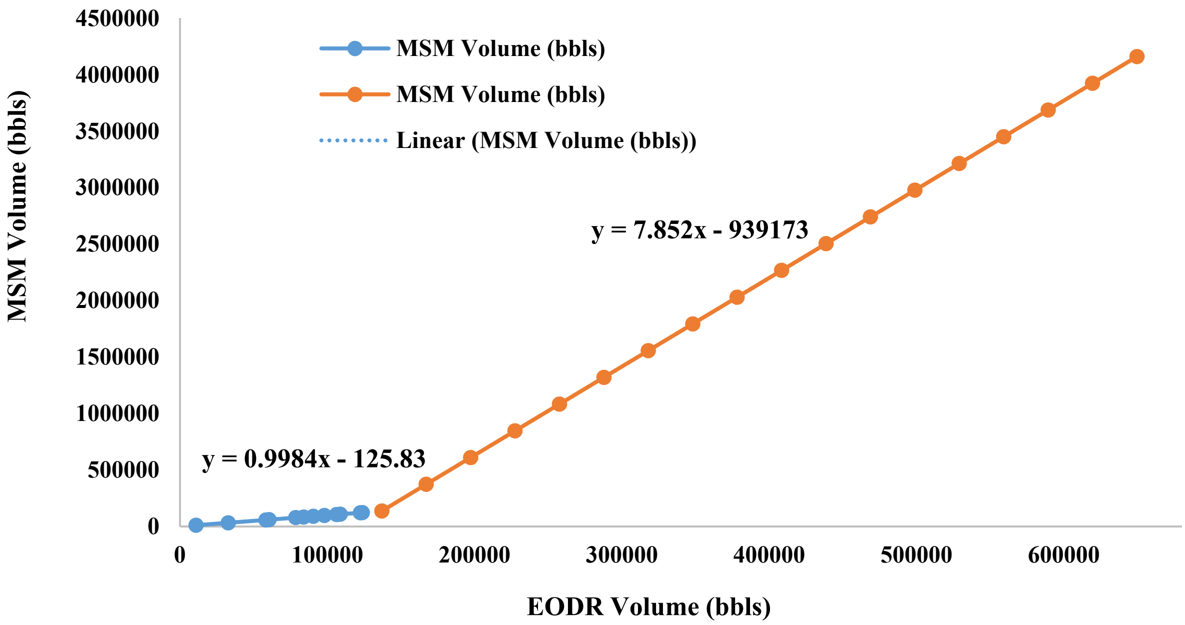

Figure 2 The graph of the MSM dataset versus the EODR dataset

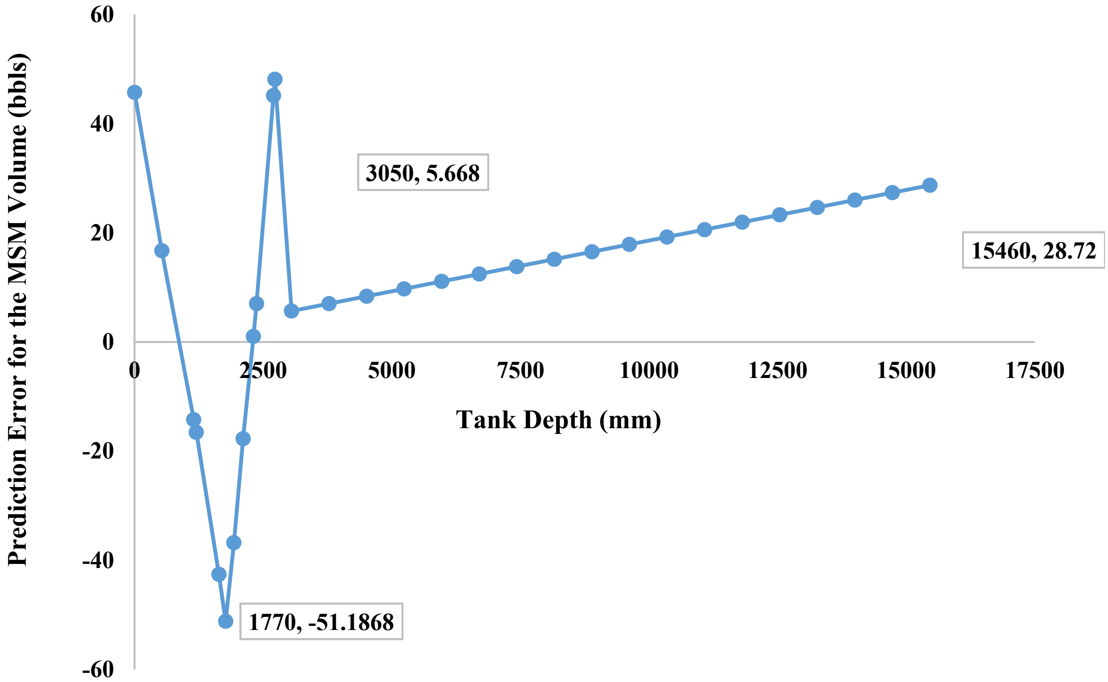

Figure 3 The graph of the prediction error versus tank depth

n_1 is the number of data records in the range of 0≤ di≤ 1770

$V_{1\,MSM(x)}$ is the MSM data records for the range of 0≤ di≤ 1770

$V_{1\,\mathrm{EODR}(x)}$ is the EODR data records for the range of 0≤ di≤ 1770

$V_{p1\,MSM(x)}$ is the predicted MSM data records for the range of 0≤ di≤ 1770

n_2 is the number of data records in the range of 1770 < di≤ 3050

$V_{2\,MSM(x)}$ is the MSM data records for the range of 1770 < di≤ 3050

$V_{2\,\mathrm{EODR}(x)}$ is the EODR data records for the range of 1770 < di≤ 3050

$V_{p2\,MSM(x)}$ is the predicted MSM data records for the range of 1770 < di≤ 3050

n_3 is the number of data records in the range of 1770 < di≤ 15460

$V_{3\,MSM(x)}$ is the MSM data records for the range of 1770 < di≤ 15460

$V_{3\,\mathrm{EODR}(x)}$ is the EODR data records for the range of 1770 < di≤ 15460

$V_{p3\,MSM(x)}$ is the predicted MSM data records for the range of 1770 < di≤ 15460

Prediction of MSM data records using the EODR data records

$$V_{p1\,MSM(x)} = (a_1)\,V_{1\,\mathrm{EODR}(x)} + c_1 \;\; for\; x = 1, 2, 3, ..n_1 \quad (8)$$

$$V_{p2\,MSM(x)} = (a_2)\,V_{1\,\mathrm{EODR}(x)} + c_2 \;\; for\; x = n_1 + 1, n_1 + 2, ..n_2 \quad (9)$$

$$V_{p3\,MSM(x)} = (a_3)\,V_{1\,\mathrm{EODR}(x)} + c_3 \;\; for\; x = n_2 + 1, n_2 + 2, ..n_3 \quad (10)$$

Where

$$n_1 + n_2 + n_3 = n \ and \ 1 \ \leq n_1 \leq n_2 \leq \ n_3 \tag{11}$$

Prediction of the MSM data records prediction error using the common tank depth, d_x data records.

$$e_{p1MSM(x)} = V_{p1\,MSM(x)} - V_{1\,MSM(x)} \tag{12}$$

$$e_{p1\,MSM(x)} = (b_1)\, d_x + \ g_1 \ for \ x = 1,2,3,..n_1 \tag{13}$$

$$e_{p2\,MSM(x)} = V_{p2MSM(x)} - V_{2\,MSM(x)} \tag{14}$$

$$e_{p2\,MSM(x)} = (b_2)\, d_x + \ g_2 \ for \ x = 1,2,3,..n_1 \tag{15}$$

$$e_{p3\,MSM(x)} = V_{p3MSM(x)} - V_{3\,MSM(x)} \tag{16}$$

$$e_{p3\,MSM(x)} = (b_3)\, d_x + \ g_3 \ for \ x = 1,2,3,..n_1 \tag{17}$$

Now

$$V_{p(K)\,MSM(x)} = V_{(K)\,MSM(x)} + e_{p(K)\,MSM(x)} \tag{18}$$

Hence

$$V_{(K)\,MSM(x)} = V_{p(K)\,MSM(x)} - e_{p(K)\,MSM(x)} \tag{18}$$

So,

$$V_{1\,MSM(x)} = V_{p1\,MSM(x)} - e_{p1MSM(x)} \tag{20}$$

$$V_{1\,MSM(x)} = (a_1)\, V_{1\,\mathrm{EODR}(x)} + \ c_1 - (b_1)\, d_x - \ g_1 \tag{21}$$

$$V_{2\,MSM(x)} = (a_2)\, V_{1\,\mathrm{EODR}(x)} + \ c_2 - (b_2)\, d_x - \ g_2 \tag{22}$$

$$V_{3\,MSM(x)} = (a_3)\, V_{1\,\mathrm{EODR}(x)} + \ c_3 - (b_3)\, d_x - \ g_3 \tag{23}$$

$$V_{1\,MSM(x)} = (a_1)\, V_{1\,\mathrm{EODR}(x)} - \ (b_1)\, d_x + \ (c_1 \ - \ g_1) \tag{24}$$

$$V_{2\,MSM(x)} = (a_2)\, V_{1\,\mathrm{EODR}(x)} - \ (b_2)\, d_x + \ (c_2 \ - \ g_2) \tag{25}$$

$$V_{3\,MSM(x)} = (a_3)\, V_{1\,\mathrm{EODR}(x)} - (b_3)\, d_x + \ (c_3 - \ g_3\,) \tag{26}$$

Essentially, the predicted MSM volume from the d_x and EODR$(x)$is defined as where $V_{MSM(x)}$

$$V_{MSM(x)} \begin{cases} (a_1)\, V_{1\,\mathrm{EODR}(x)} - \ (b_1)\, d_x + \ (c_1 \ - \ g_1)\, for \ x = 1 \ to \ n_1 \\ (a_2)\, V_{1\,\mathrm{EODR}(x)} - \ (b_2)\, d_x + \ (c_2 \ - \ g_2)\, for \ x = n_1 + 1 \ to \ n_2 \\ (a_3)\, V_{1\,\mathrm{EODR}(x)} - (b_3)\, d_x + (c_3 - \ g_3) for \ x = n_2 + 1 \ to \ n_3 \end{cases} \tag{27}$$

3. Results and discussion

3.1 The results of the Linear Regression (LR) Model

The hyperparameters of the Linear Regression (LR) model is presented in Table 1. Also, the results for the LM model is presented in Figure 4 for residual plot, Figure 5 for the volume versus tank depth plot, Figure 5 for the plot of MSE, RMSE, MAE, Figure 7 for the confusion matrix and R^2 versus epoch , Table 2 for the error metric scores versus epoch and Table 3 for the overall or steady state metric scores of the LM model.

The results as presented in Table 3 showed that the LM model has MSE of 0.1997, RMSE of 0.4469, MAE of 0.4201 and prediction accuracy expressed as 1-MAPE having a value of 0.9999%. Also, the LM model has the minimum error of 0.0172 at the tank depth 1880.0 and the maximum error of 13.8777 at the tank depth 8210.0., as shown in Table 3.

Table 1: Hyperparameters of the Linear Regression (LR) Model

Hyperparameter	Value	Description
Polynomial degree	2	The model considers both linear and quadratic relationships between Depth_mm and Volume_L_EODR.
Regularization strength	$\alpha = 0.1$	This value is set to prevent overfitting
Epochs	5	The model is trained for 5 iterations to monitor performance improvements over time.

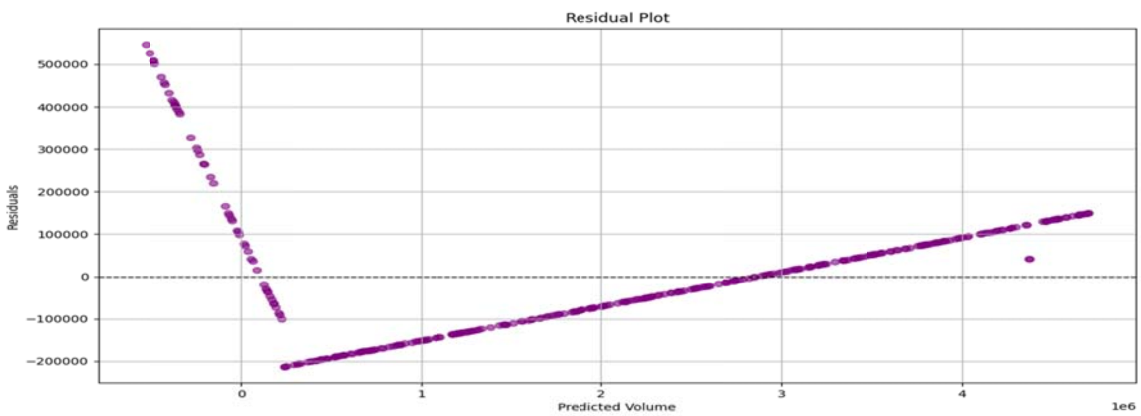

Figure 4: Residual plot for **LM model**

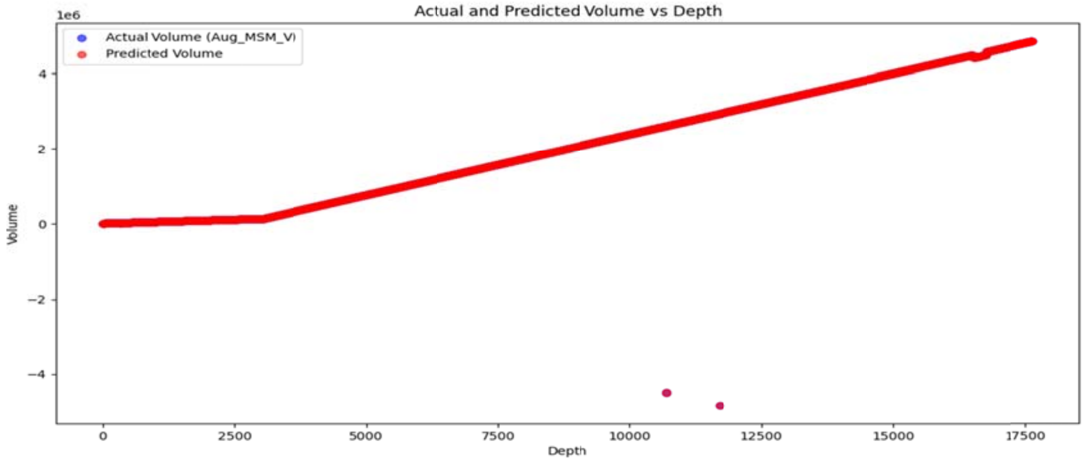

Figure 5: Volume vs Tank Depth for **LM model**

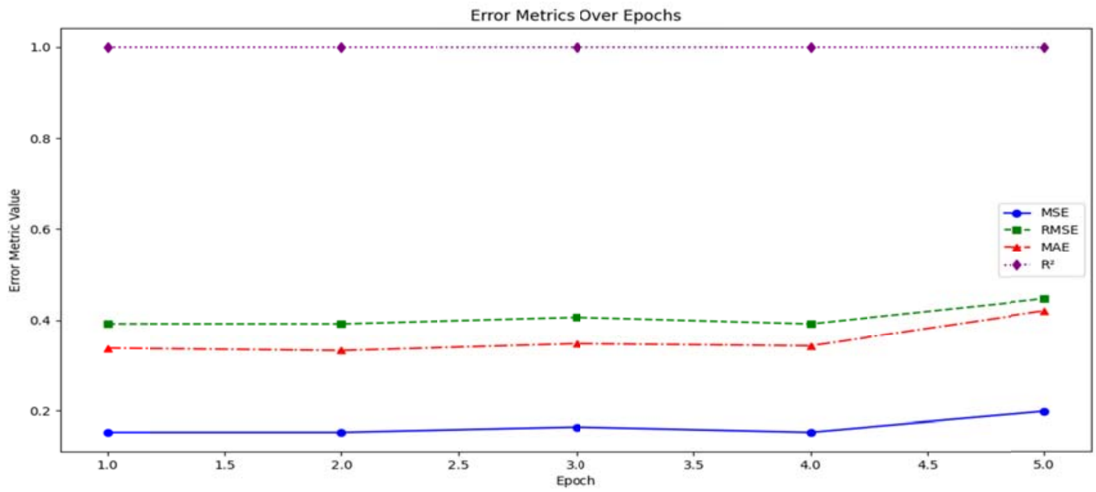

Figure 6: Error metrics for **LM model**

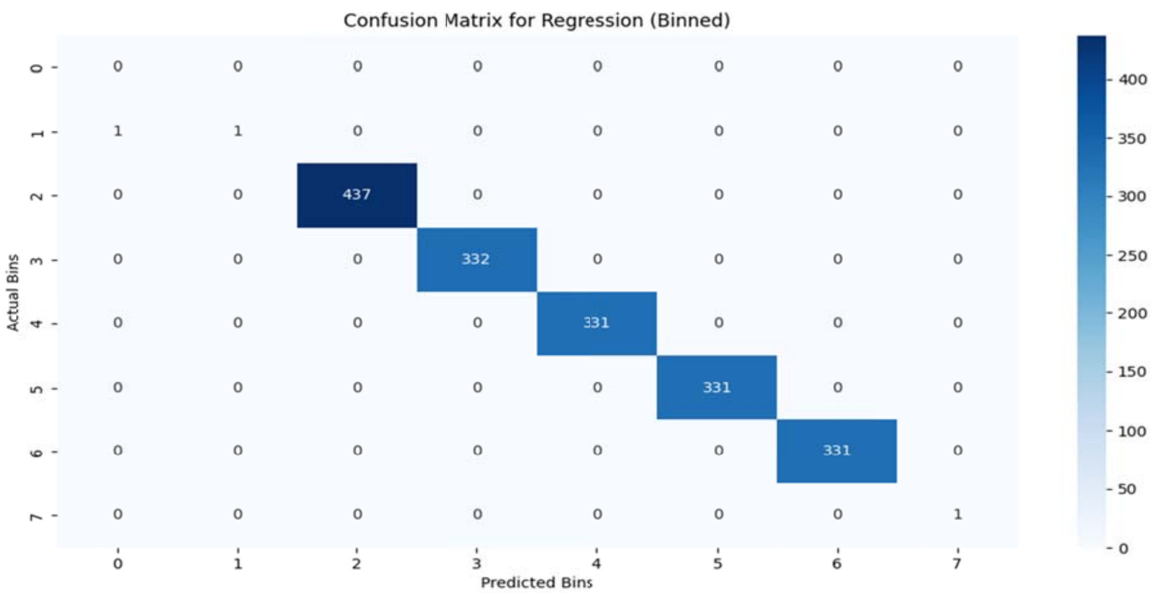

Figure 7 The Confusion matrix for **ML model**

Table 2: Error metrics score versus epoch for LM model

	Score					
Epoch	MSE for LRM	RMSE for LRM	MAE for LRM	R^2 for LRM	MAPE for LRM (%)	Percentage Error (%)
1	0.1536	0.3919	0.3384	1.0000	0.0100%	0.0100%
2	0.1536	0.3919	0.3330	1.0000	0. 0100%	0.0100%
3	0.1652	0.4064	0.3482	1.0000	0. 0100%	-0.0001%
4	0.1536	0.3919	0.3435	1.0000	0. 0100%	0.0100%
5	0.1997	0.4469	0.4201	1.0000	0. 0100%	-0.0100%

Table 3 Error metrics at the end of Epoch for LM model

Metric	Score for LRM
MSE	0.1997
RMSE	0.4469
MAE	0.4201
R^2.	0.9987
MAPE	0.0001%
Percentage error	-0.0001%
Min Error	0.0172 at Depth 1880.0
Max Error	13.8777 at Depth 8210.0

3.2 The results of the Piecewise Multiple Linear Regression Function (PMLRF) Model

The results of the Piecewise multiple linear regression function (PMLRF) model is presented in Figure 8 for the volume versus tank depth plot, Figure 9 for residual plot, and Table 4 for the overall or steady state metric scores of the PMLRF model. The results as presented in Table 4 showed that the PMLRF model MSE of 0.18770, RMSE of 0.39365, MAE of 0.36112 and prediction accuracy expressed as 1-MAPE having a value of 0.9999%. Also, the PMLRF model has the minimum error of -0.0948 at the tank depth 2100 and the maximum error of 2.5632 at the tank depth 2420, as shown in Table 4.

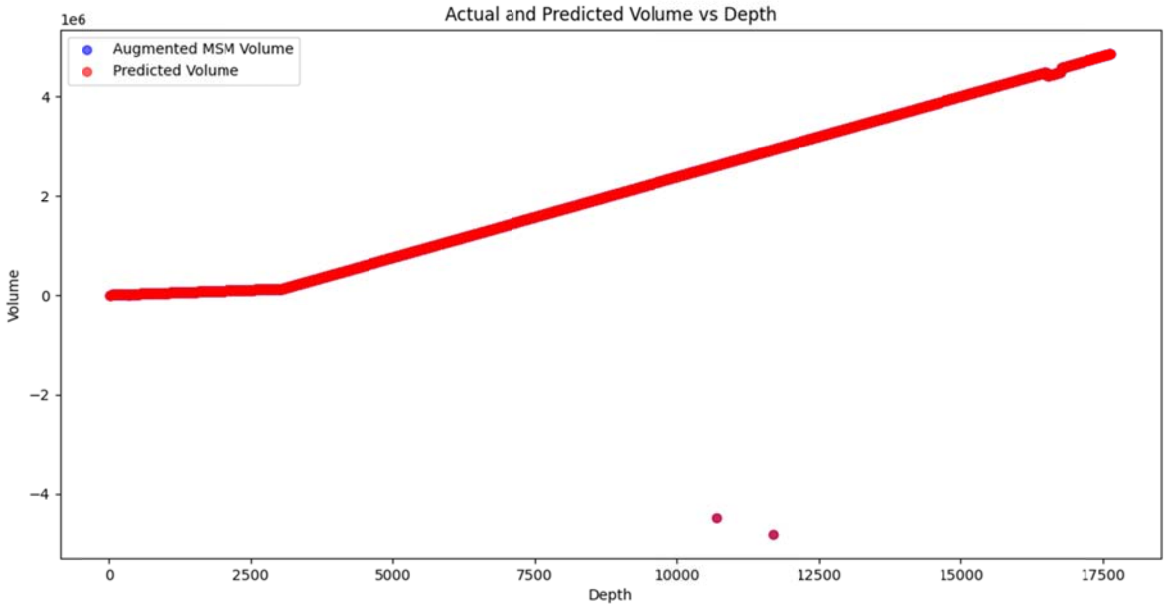

Figure 8: Actual and predicted volume vs tank depth for the Piecewise multiple linear regression function (PMLRF) Model

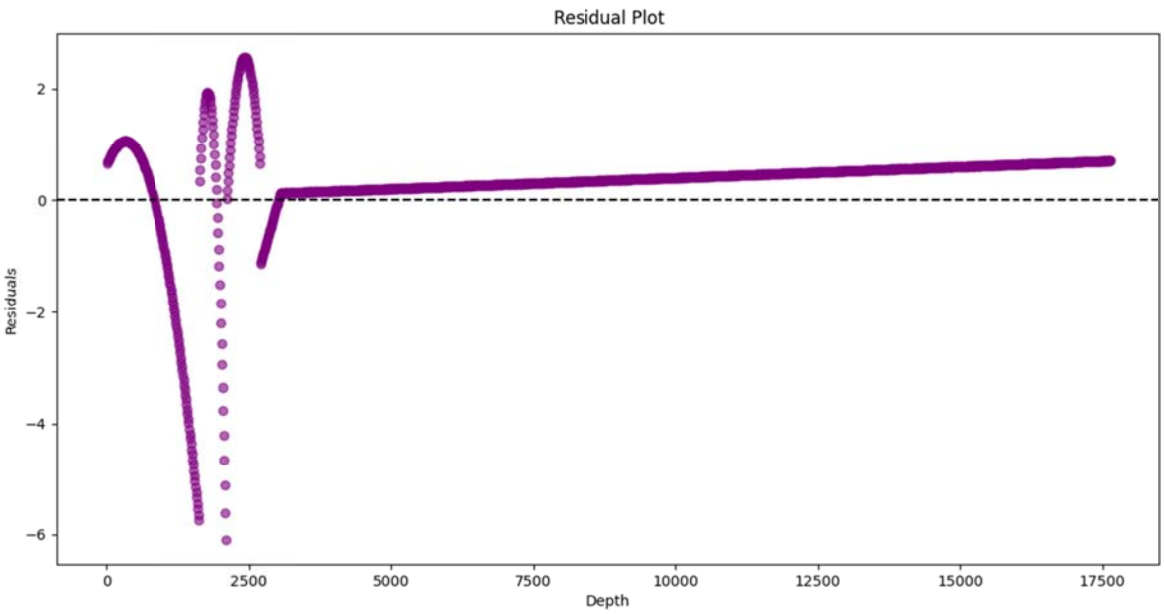

Figure 9: Residual plot for **LM model**

Table 4 Error metrics at the end of Epoch for **the Piecewise multiple linear regression function (PMLRF) Model**

Metric	Score for PMLRM
MSE	0.18770
RMSE	0.39365
MAE	0.36112
R^2.	0.99965
MAPE	0.0009%
Percentage error	-0.00004%
Min Error	-0.0948 (at Depth 2100)
Max Error	2.5632 (at Depth 2420)

4. Conclusion

This study successfully developed and validated Linear Regression (LR) and Piecewise Multiple Linear Regression Function (PMLRF) models designed to bridge the accuracy gap between Electronic Optical Distance Ranging (EODR) and the traditionally accurate Manual Strapping Method (MSM). As manual strapping is highly tedious and time-consuming, this research establishes that the developed models, particularly the piecewise approach, can accurately predict the "ground truth" MSM volume using data acquired from the more efficient EODR method.

By training on historical datasets, the proposed models minimize the high variance (error) between the two methods. Consequently, this study provides a viable, highly accurate, and time-efficient alternative for tank calibration, effectively eliminating the need for tedious manual calibration methods in modern oil storage applications. The models enable automated, precise volume prediction, ensuring that the EODR-based data meets necessary industry standards (e.g., API) without requiring direct, manual measurement of the tank.

References

1. Shunashu, I. L., & Casmir, R. (2020). Assessing the impact of measurement uncertainty in custody transfer to the development of oil & gas industry in Tanzania. *Business Education Journal*, *6*(2).

2. Potten, G. (2021, August). Realizing value from adherence to crude oil custody transfer automatic sampling standards. In *SPE Nigeria Annual International Conference and Exhibition* (p. D031S013R002). SPE.
3. Agboola, O. O., Akinnuli, B. O., Akintunde, M. A., Ikubanni, P. P., & Adeleke, A. A. (2019, December). Comparative analysis of manual strapping method (MSM) and electro-optical distance ranging (EODR) method of tank calibration. In *Journal of Physics: Conference Series* (Vol. 1378, No. 2, p. 022062). IOP Publishing.
4. Hassani, S., & Dackermann, U. (2023). A systematic review of advanced sensor technologies for non-destructive testing and structural health monitoring. *Sensors*, *23*(4), 2204.
5. Fu, Y., Shang, Y., Hu, W., Li, B., & Yu, Q. (2021). Non-contact optical dynamic measurements at different ranges: a review. *Acta Mechanica Sinica*, *37*(4), 537-553.
6. James, M. R., Robson, S., & Smith, M. W. (2017). 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys. *Earth Surface Processes and Landforms*, *42*(12), 1769-1788.
7. Adamson, B., Waskom, M., Blarre, A., Kelly, J., Krismer, K., Nemeth, S., ... & Cohen, A. B. (2023). Approach to machine learning for extraction of real-world data variables from electronic health records. *Frontiers in Pharmacology*, *14*, 1180962.
8. Agboola, O. O., Akinnuli, O. B., Akintunde, A. M., & Kareem, B. (2020). Modelling of cost estimates for the geometrical calibration of upright oil storage tanks. *International Journal of Energy Economics and Policy*, *10*(1), 464-470.
9. Agboola, O. O., Akinnuli, B. O., Akintunde, M. A., Ikubanni, P. P., & Adeleke, A. A. (2019, December). Comparative analysis of manual strapping method (MSM) and electro-optical distance ranging (EODR) method of tank calibration. In *Journal of Physics: Conference Series* (Vol. 1378, No. 2, p. 022062). IOP Publishing.
10. Tariq, Z., Aljawad, M. S., Hasan, A., Murtaza, M., Mohammed, E., El-Husseiny, A., ... & Abdulraheem, A. (2021). A systematic review of data science and machine learning applications to

the oil and gas industry. *Journal of Petroleum Exploration and Production Technology*, *11*(12), 4339-4374.

11. Ibrahim, N. M., Alharbi, A. A., Alzahrani, T. A., Abdulkarim, A. M., Alessa, I. A., Hameed, A. M., ... & Almuqhim, A. A. (2022). Well performance classification and prediction: Deep learning and machine learning long term regression experiments on oil, gas, and water production. *Sensors*, *22*(14), 5326.
12. Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, *6*(4), 379-391.