

Data Augmentation For Crude Oil Storage Tank Calibration Dataset Using Piecewise Non-Linear Function

Emmanuel Udama Odeh¹

Department of Mechanical and Aerospace Engineering,
University of Uyo, Uyo Akwa Ibom State Nigeria
emmanuelodeh@uniuyo.edu.ng

John Dennis Urua²

Department of Mechanical and Aerospace Engineering,
University of Uyo, Uyo Akwa Ibom State Nigeria
johnurua1@gmail.com

Aaron Tukpakonbowei Ben-Elijah³

Ralph E. Martin Department of Chemical Engineering,
University of Arkansas, USA
aaron.benelijah@gmail.com
ORCID No: 0009-0009-0245-1650

Abstract—This research presents a data-driven approach to enhance the accuracy of crude oil storage tank calibration by augmenting Electro-optical Distance Ranging (EODR) data using Manual Tank Strapping Method (MSM) data as the ground truth. To improve dataset resolution for automatic inventory computation, a piecewise non-linear function (PNLF) was applied to field-measured data from an oil production facility. The PNLF method effectively captures the non-linear, irregular characteristics of tank structures and sensor readings. Results indicate that the PNLF method closely tracks actual MSM volumes, with a low maximum absolute error of 0.0979 bbls. Statistical validation confirms high-fidelity calibration, with a Mean Absolute Error (MAE) of 0.03009333 bbls and a Root Mean Squared Error (RMSE) of 0.0365285 bbls, demonstrating minimal deviation relative to typical tank volumes. This approach proves superior for improving the reliability of automated crude oil inventory management.

Keywords—Crude Oil Storage Tank Calibration, Data Augmentation, Machine Learning Model, Piecewise Non-Linear Function

1. Introduction

In the global oil and gas industry, crude oil storage tanks serve as the primary units for fiscal accounting and inventory management [1,2]. Minor inaccuracies in tank calibration, the process of determining a tank's precise volume at various liquid heights, can lead to significant financial discrepancies during custody transfer and production reporting [3,4]. Notably, effective crude oil inventory management depends fundamentally on the precision of storage tank calibration. While the Manual Tank Strapping Method

(MSM) has long been the primary benchmark for ground truth measurements, its labor-intensive nature and limited data resolution make it less efficient than contemporary digital alternatives [5,6].

The emergence of Electro-optical Distance Ranging (EODR), utilizing laser-based scanners, has introduced high-resolution spatial mapping to the calibration process [7]. Despite its precision, EODR data can be influenced by environmental factors and instrument noise, necessitating a robust alignment with established physical standards [8,9]. Furthermore, structural irregularities such as shell deformations, welding seams, and construction variances create non-uniformity in tank radii, which linear interpolation models often fail to capture accurately [10].

This study addresses these limitations by implementing a data-driven augmentation strategy. By leveraging a piecewise non-linear function, the research aims to bridge the gap between high-resolution EODR datasets and the physical reliability of MSM data. This approach optimizes the dataset resolution for automatic inventory computation, ensuring that corrections for hydrostatic head, temperature, and tilt are integrated into a more precise, augmented model of the tank's internal geometry.

2. Methodology

This research employs a data-driven approach to enhance the accuracy of crude oil storage tank calibration by augmenting Electro-optical Distance Ranging (EODR) data using Manual Tank Strapping Method (MSM) data as the ground truth. The study utilizes field-measured data from an oil production field, processed through a piecewise non-linear

function to improve dataset resolution for automatic inventory computation.

2.1 Data collection and preprocessing

Two primary datasets were extracted from the certified calibration reports of vertical crude oil storage tanks. The first dataset is the Manual Tank Strapping Method (MSM) Data. This is considered the ground truth in this study, and it represents the physical measurements of the tank shell. The second dataset is the Electro-optical Distance Ranging (EODR) Data. This is the high-resolution digital measurements obtained via a laser-based total station scanner, compliant with industry standards.

The raw datasets were processed according to industry standards. In the MSM data processing, the internal radii were calculated, and capacity tables developed in accordance with the American Petroleum Institute (API) Manual of Petroleum Measurement Standards (MPMS). The following corrections were applied to both datasets to ensure accuracy: correction for hydrostatic head effect; correction to certified tank shell temperature and correction for deadwood and tilt.

The EODR data was mapped to match MSM measurement levels, using liquid height as the alignment variable. This approach treats MSM as the ground truth, enabling a standardized spatial analysis of the internal tank radius across both datasets.

2.2 Data Augmentation Procedure (Piecewise Non-Linear Function)

Due to the non-uniformity in tank shell radius, caused by deformation, welding, or construction irregularities, a simple linear interpolation is insufficient. Therefore, a piecewise non-linear function is applied. The Piecewise Non-Linear Function (PNLF) for data augmentation approach is used. This involves segmented modeling, piecewise function application and augmentation generation. In the segmented modeling, the tank shell is divided into distinct segments (e.g., based on shell courses or specific height intervals). In the piecewise function application, a piecewise non-linear function is fitted to the EODR data within each segment to model the localized, non-linear profile of the tank. In the data augmentation, the piecewise function is used to generate intermediate data points (example, increasing measurement resolution from every 1 meter to every 0.1 meter) to create a denser, more accurate calibration table.

Specifically, the MSM volume is determined using six different non-linear expressions presented as a piecewise non-linear function. In each of the six different non-linear expressions, the MSM volume is expression as function of EODR volume and tank depth which is defined as follows;

$$M_{P(K)} = a_x(E_{(K)}) + b_x(D_{(K)})^2 + c_x(D_{(K)}) + d_x \quad (1)$$

$$M_{P(K)} = M_{a(K)} + e_{(K)} \quad (2)$$

$$e_{(K)} = M_{P(K)} - M_{a(K)} \quad (3)$$

$$e_{(K)} = a_x(E_{(K)}) + b_x(D_{(K)})^2 + c_x(D_{(K)}) + d_x - M_{a(K)} \quad (4)$$

where;

$E_{(K)}$ denote the EODR volume at data point k

$D_{(K)}$ denote the tank depth at data point k

$M_{a(K)}$ denote the actual MSM volume at data point k

$M_{P(K)}$ denote the predicted MSM volume at data point k

$e_{(K)}$ denote the prediction error at data point k

X denote the non-linear function number

The values of the model coefficients a,b,c, and d are given for each of the function. Also, the value of x (the function number) corresponds to the row number in used to provide the values of the model coefficients a,b,c, and d. The value of x is used to match the coefficients to the functions. Accordingly, the expression for the piecewise non-linear function is given as follows;

$$M_{P(K)} = \begin{cases} a_1(E_{(K)}) + b_1(D_{(K)})^2 + c_1(D_{(K)}) + d_1 ; 0 < D_{(K)} < 1640 \\ a_2(E_{(K)}) + b_2(D_{(K)})^2 + c_2(D_{(K)}) + d_2 ; 1640 < D_{(K)} < 1930 \\ a_3(E_{(K)}) + b_3(D_{(K)})^2 + c_3(D_{(K)}) + d_3 ; 1930 < D_{(K)} < 2110 \\ a_4(E_{(K)}) + b_4(D_{(K)})^2 + c_4(D_{(K)}) + d_4 ; 2110 < D_{(K)} < 2700 \\ a_5(E_{(K)}) + b_5(D_{(K)})^2 + c_5(D_{(K)}) + d_5 ; 2700 < D_{(K)} < 3780 \\ a_6(E_{(K)}) + b_6(D_{(K)})^2 + c_6(D_{(K)}) + d_6 ; 3780 < D_{(K)} < 17641 \end{cases} \quad (5)$$

2.3 The Procedure for Data Augmentation using Piecewise Non-Linear Function (PNLF)

This procedure outlines the data augmentation technique for improving crude oil storage tank calibration tables by generating high-resolution (denser) calibration data from lower-resolution Electro-Optical Distance Ranging (EODR) measurements. It uses a segmented, non-linear modeling approach to handle deformations, welding irregularities, and non-uniform shell radii.

Step1 : Data Preparation and Preprocessing

Step 1.1 Input Data: Gather the initial EODR dataset consisting of tank depths ($D_{(k)}$) and corresponding EODR-measured volumes ($E_{(k)}$).

Step 1.2 Actual Data: Obtain the Manual Strapping Method (MSM) volumes ($M_{a(k)}$) at specific, sparse intervals (such as, every 1 meter) to act as calibration checkpoints.

Step 1.3 Identify Deformation Points: Detect areas of significant tank deformation or welding irregularities where linear interpolation fails.

Step 2 Segmented Modeling (Segmentation)

Step 2.1 Divide the entire vertical height of the storage tank into N distinct segments ($S_{(1)}, S_{(2)}, \dots, S_{(n)}$) based on shell courses, height intervals, or identified deformation zones.

Step 2.2 Within each segment, identify the localized non-linear profile.

Step 3 Piecewise Function Application (Fitting)

Step 3.1 Within each segment, apply a piecewise non-linear function to model the relationship between EODR volumes and MSM volumes.

Step 3.2 For each segment, select and fit one of the six specific non-linear expressions (represented by index x) to the EODR and MSM data points to compute the model coefficients $a_x, b_x, c_x,$ and d_x .

Step 3.3 The predicted MSM volume ($M_{p(k)}$) at a given point K is calculated using Equation 1.

Step 4 Error Analysis and Model Refinement

Step 4.1 Calculate the prediction error (ϵ) for each segment using the fitted function and the actual MSM data using Equation 4.

Step 5 Data Augmentation Generation (Interpolation)

Step 5.1 Increase the measurement resolution within each segment (such as from 1-meter intervals to 0.1-meter intervals).

Step 5.2 Generate intermediate EODR data points within the segments.

Step 5.3 Apply the fitted piecewise non-linear function to these new, finer-interval data points to calculate corresponding, high-density, accurate MSM volumes.

Step 5.4 Merge the original data points with the newly generated points to create a comprehensive, dense, and non-linear, high-accuracy calibration table.

Step 6 Validation

Step 6.1 Compare the augmented dataset against independent validation measurements (if available) to ensure the non-linear interpolation reduces, rather than introduces, calibration errors.

3. Results and discussion

4.2.2 Results of the Data Augmentation using the Piecewise Non-Linear Function (PNLF)

The PNLF method demonstrates high accuracy in forecasting MSM volume (bbls), as shown by the close alignment of actual and predicted values in Figure 1. Figure 2 confirms this, highlighting a low absolute error of 0.0979 bbls. This performance is consistent with the statistical metrics achieved via the Gaussian method on the augmented dataset (Table 1).

The line chart (Figure 1) comparing actual and predicted MSM volumes (bbls) indicates that the PNLF method closely tracks the actual values. This implies that the piecewise approach effectively captures the non-linear characteristics of the tank calibration data, which often arise from irregularities in tank structure or non-linear sensor readings. Again, the scatter plot of the prediction error (Figure 2) demonstrates a high level of accuracy. An absolute error of approximately 0.0979 bbls is reported, which indicates that the PNLF model generates, on average, a very low divergence from the target volume, ensuring high-fidelity calibration.

The Mean Absolute Error (MAE) of 0.03009333 bbls and Root Mean Squared Error (RMSE) of 0.0365285 bbls show that the average error is very small relative to the typical total volume of a crude oil storage tank (often in the range of hundreds of thousands of barrels).

Again, the scatter plot reveals a maximum deviation of only 0.0979 bbls. Considering typical custody transfer tanks hold vast amounts (such as 600,000 bbls), an error of less than 0.1 barrels represents extremely high accuracy. Also, the lowest observed error is nearly zero, highlighting the stability of the PNLF method across the data points. Again, the PNLF method, combined with data augmentation, provides a reliable technique for crude oil storage tank calibration, delivering a low absolute error (0.0979 bbls) and demonstrating strong performance in modeling non-linear volumetric relationships.

In all, the PNLF method effectively models the non-linear relationship in the tank calibration data, successfully increasing dataset size without sacrificing accuracy. The very high R^2 and extremely low maximum error make it a suitable method for generating accurate, reliable data for tank capacity tables.

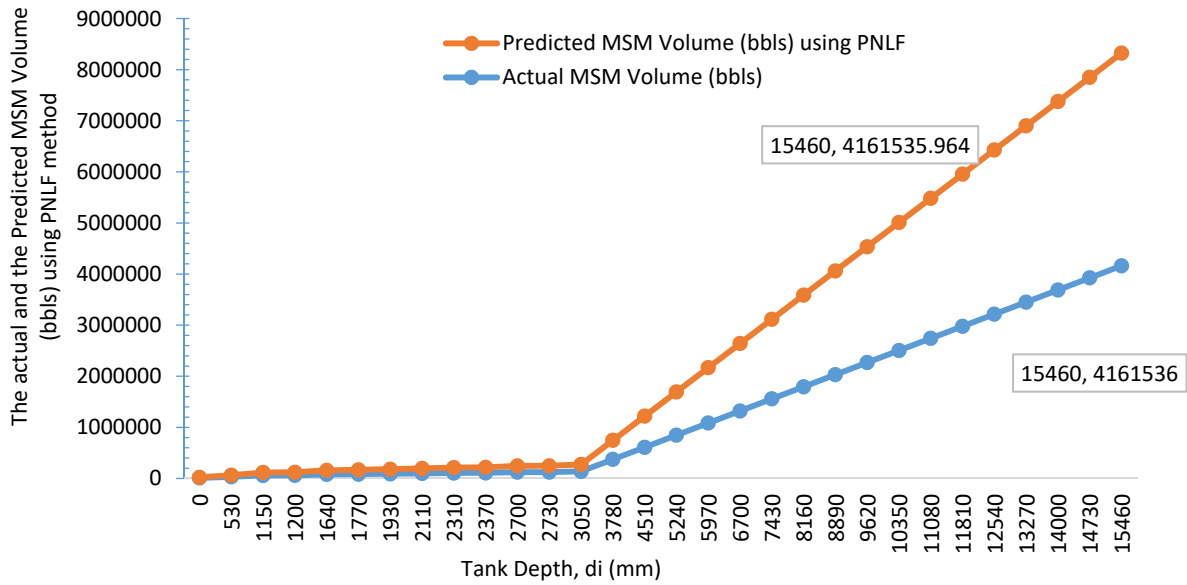


Figure 1: The line chart of the actual and the predicted MSM volume (bbls) using the Piecewise Non-Linear Function (PNLF) method

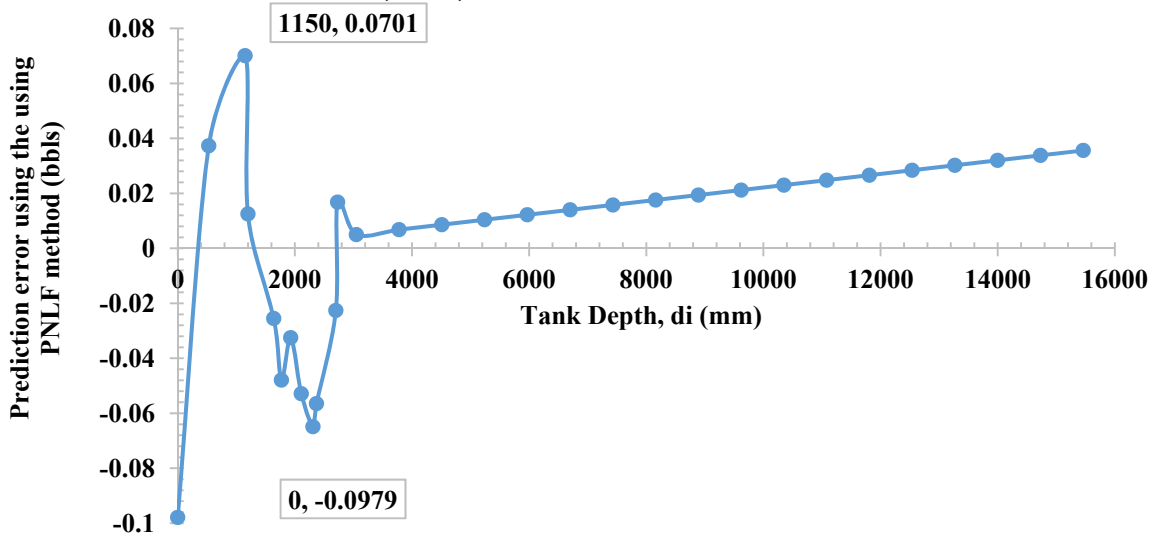


Figure 2: The scatter plot of the prediction error of the MSM volume (bbls) using the Piecewise Non-Linear Function (PNLF) method

Table 1: Statistical metrics for the augmented dataset using the Piecewise Non-Linear Function (PNLF) method

	Prediction performance of the PNLF method
MSE	0.00133433
RMSE	0.0365285
MAE	0.03009333
R ²	1.00000
1-MAPE	0.99996883
Mean of Absolute Percentage Error (%)	0.000051
Minimum Absolute Error	0.005
Maximum Absolute Error	0.0979

4. Conclusion

The study validates a data-centric framework for enhancing crude oil tank calibration. It utilizes piecewise non-linear augmentation to bridge the gap between EODR measurements and manual strapping (MSM) ground truths, specifically targeting accuracy losses caused by welding artifacts and tank shell deformations. Notably, the methodology effectively aligned high-resolution EODR scanner data with MSM physical measurements by applying rigorous API MPMS corrections, including hydrostatic head, temperature, and tilt adjustments. Also, the implementation of the piecewise non-linear function allowed for the successful modeling of non-uniform tank radii, providing a more granular dataset than traditional strapping methods alone. Again, the augmentation technique facilitates more precise automatic inventory computation, reducing the margin of error in volume calculations for oil production fields. In all, the integration of piecewise modeling with hybrid measurement datasets provides a robust framework for digital tank calibration. This approach not only upholds industry standards but also enhances the reliability of digital twin models used in modern oil storage management.

References

1. Ali, U., Salah, B., Naeem, K., Khan, A. S., Khan, R., Pruncu, C. I., & Khan, S. (2020). Improved MRO inventory management system in oil and gas company: Increased service level and reduced average inventory investment. *Sustainability*, 12(19), 8027.
2. Wokeh, P. I., & Nmehielle, C. O. (2023). Inventory Management Techniques and Financial Performance Of Listed Oil And Gas Companies In Nigeria. *BW Academic Journal*, 16-16.
3. Patel, S., Parrott, B., Abdellatif, F., & Trigui, H. (2020, November). Custody Transfer Tank Calibration Technology. In *Abu Dhabi International Petroleum Exhibition and Conference* (p. D012S116R012). SPE.
4. 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).
5. 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.
6. Marques, M., Sousa, J. A., & Ribeiro, L. (2019). Calibration 4.0—Information system for usage of digital calibration certificates. In *19th International Congress of Metrology (CIM2019)* (p. 01002). EDP Sciences.
7. National Research Council, Division on Engineering, Physical Sciences, & Committee on Review of Advancements in Active Electro-Optical Systems to Avoid Technological Surprise Adverse to US National Security. (2014). *Laser radar: progress and opportunities in active electro-optical sensing*. National Academies Press.
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. Wang, J. T., Liu, Z. Y., Zhang, L., Guo, L. G., Bao, X. S., & Tong, L. (2010). Automatic measurement system for vertical tank volume by electro-optical distance-ranging method. *Applied Mechanics and Materials*, 26, 416-421.
10. Spritzer, J. M., Bohra, H., & Guzey, S. (2019). Imperfection-sensitivity of unanchored aboveground open-top steel welded liquid storage tanks subjected to seismic loads. *SN Applied Sciences*, 1(12), 1566.