J. Watada et al. (Eds.)

doi:10.3233/AERD230005

Reconstruction of Flow Rate History Using Linear Regression

Berihun Mamo NEGASH[1], Chee Him POON, Pandian M VASANT

[1] Institute of Hydrocarbon Recovery and Petroleum Engineering Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia.
ORCiD ID: Berihun Mamo Negash https://orcid.org/0000-0002-2208-8504

Abstract. Long-term pressure and flow rate history are important for reservoir characterization and reservoir management. However, a complete set of these data are often not available due to numerous technical difficulties. Currently, datasets with missing information are omitted and not considered for further analysis. In this study, we use machine learning algorithm via linear regression for flow rate history reconstruction. Only few studies have demonstrated the application of linear regression for well testing purposes. However, pressure and flow rate data in a producing field are comparatively longer and more complex. A combination of feature extraction and linear regression was applied for long term flow rate history reconstruction. The dataset used to evaluate the performance of the proposed method was obtained from a real producing field. This study indicated the high performance of linear regression at estimating missing flow rate history using available pressure readings in the dataset. Although linear regression has the benefits of high interpretability and fast computation time, it fails to perform well in reconstructing flow rate history when there is a significant degree of variation in the flow rate and pressure data.

Keywords: Flow rate history reconstruction · Reservoir management · Linear regression · Machine learning · Statistical learning

1 Introduction

Flow rate and pressure response history are certainly important as a rich source of information about the reservoir. Having access to a complete set of flow rate and pressure response history is of paramount importance for reservoir characterization, behavior understanding, and future performance prediction. The reconstruction of flow rates and pressure responses over time provides valuable insights into the reservoir's dynamics, allowing engineers to identify and account for subsurface features such as fractures or faults that may affect fluid flow. These insights are critical for optimizing reservoir performance and maximizing economic potential. Accurate predictions about reservoir behavior under different production scenarios can inform important decisions about drilling, production strategies, and reservoir management. Therefore, the availability of comprehensive flow rate and pressure response history is essential for the successful development and management of reservoirs. However, obtaining a complete history of production data with its respective pressure response is especially difficult to achieve due to

[1] Corresponding author: Berihun Mamo Negash, berihunn@gmail.com

number of reasons. Firstly, pressure record is subjected to dynamic changes such as sudden changes in flow temperature, that might occur in the wellbore or the reservoir [1]. Secondly, flow rate data is usually measured and recorded on the surface in varying time intervals, which result in incomplete flow rate history with irregular time interval between them [2].

Missing flow rate history could be calculated using conventional model-based approach with the assumption that there are negligible changes on the reservoir parameters over the production time. However, the process can be tedious and time-consuming. To effectively reconstruct the flow rate history for the reservoir, machine learning emerged as an attractive alternative that is advantageous in terms of fast computational time and less susceptible to bias than human interpretation [3]. By utilizing the available data, machine learning algorithms 'learn' the underlying relationships between the datasets. As a result, data-driven model can generalize more realistically to the reservoir compared to model-driven solution. In fact, The utilization of machine learning are extended to wide range of application in petroleum engineering, such as production forecasting [3-5], optimization of surface facilities [6], history matching [7, 8], well parameter monitoring [1, 9], speed up reservoir simulation [10] and etc. Typical data collected from a producing well, such as bottom-hole pressure and flow rate can be utilized for the purposes. In fact, several studies applying machine learning techniques on data collected from PDG have been done to simulate well testing and flow rate reconstruction [9, 11, 12]. However, the previous studies were done in the context of well test analysis, where the timeframe is relative short compared to field production time. Besides, the variation presented in the pressure and flow rate data was relatively small. Therefore, the linear regression algorithm might not capable of predicting missing flow rate history in long-term dataset that generally incorporates large degree of fluctuation. In this study, a combination of feature extraction and linear regression algorithm will be applied to flow rate history reconstruction upon long-term pressure and flow rate dataset. The dataset used was collected from a real producing field over eight years. The effectiveness of linear regression on flow rate history reconstruction will be assessed using two datasets, one with small degree of variation and another with high degree of variation in pressure and flow rate data.

2 Literature Review

Machine learning intersects with areas of engineering, statistics, data mining and artificial intelligence. The development of predictive model for production forecast is classified as a supervised learning problem, where both the output y and features x are measured, and the goal is to find the pattern behind y and x, and used the trained pattern to make prediction [13]. Within the category of supervised learning, reconstruction of flow rate history is classified as a regression problem, where its output is essentially quantitative (e.g. flow rate). Thus, machine learning technique for regression will be applied in the study of flow rate history reconstruction.

Numerous researches has been conducted to interpret flow rate and pressure data using machine learning approaches for different applications. Notably, studies about the interpretation of pressure and flow rate data collected from permanent downhole gauge (PDG) has been done by researchers for well test analysis using machine learning [9, 11, 12, 14]. The results of the studies demonstrated that machine learning is capable of extracting relationship between the parameters in the data. Relationship between flow rate and pressure was expressed in the form of pressure convolution, then applied to well testing dataset [9]. Study on flow rate reconstruction was also conducted on different reservoir models such as homogenous infinite reservoir, channel reservoir and reservoir near a sealing fault [1].

However, algorithms introduced in the previous studies were verified using synthetic dataset and simple real dataset in the context of well test analysis. The dataset has a minimal degree of variation; hence it is relatively simpler to model. This study focused on flow rate history reconstruction upon long-term real dataset by extending the algorithm proposed by the previous studies. The timeframe of the real dataset is comparatively longer. Besides, the challenges of flow rate history reconstruction on long-term real dataset are the large number of uncertainties involved and the fluctuation of measured data. Therefore, the combination of feature extraction and linear regression algorithm in this study will be tested to assess its robustness in modeling complex flow rate and pressure dataset over a long timeframe.

3 Problem Statement

The traditional methods of flow rate history reconstruction, such as analytical solutions and reservoir simulations, require extensive investments in terms of time, money, and effort. As a result, there is a need for a more efficient and cost-effective approach to reconstructing flow rate history. In response to this need, this study aims to develop a data-driven predictive model using linear regression for the reconstruction of flow rate history on real datasets. The performance of the linear regression approach is evaluated using both simple synthetic and complex real datasets. However, the limitations and assumptions of linear regression must be carefully considered, especially in handling complex and non-linear relationships between variables, and the model's performance must be validated against existing analytical solutions or simulation results to ensure reliable and accurate predictions. The overall goal of this study is to offer a viable alternative to traditional methods that can optimize reservoir performance and maximize economic potential.

4 Methodology

Linear Regression

Linear regression has long been a topic of interest in various industries [15-19]. Linear regression is a statistical learning method that models the relationship between a dependent variable, y and one or multiple independent variables, x. Linear regression is a computationally fast algorithm with high interpretability through weightage of each feature on the output. Each term within the features reflects the physical properties of flow rate in relation with pressure and time [11]. The features were formulated by mapping the flow rate based on pressure convolution was shown in Eq.1 [9].

$$x^{(i)} = \begin{bmatrix} \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)}) \\ \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)}) \log(t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)})(t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)}) / (t^{(i)} - t^{(j)}) \\ t^{(i)} \end{bmatrix}, i = 1, \ldots, m \tag{1}$$

In the case where both bottomhole pressure and wellhead pressure were collected, both pressure responses were being utilized to extract their relationships with the output. Since wellhead pressure is strongly correlated with bottomhole pressure, the wellhead pressure will implicitly contains information about the reservoir flow. Therefore, features were mapped using both bottomhole pressure and wellhead to capture more information about the reservoir. Some additional features were introduced to model the ratio between bottomhole pressure and wellhead pressure, and the differential pressure between bottomhole pressure and wellhead pressure.

$$x^{(i)} = \begin{bmatrix} \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)}) \\ \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)}) \log(t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)})(t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{bhp}{}^{(j)} - P_{bhp}{}^{(j-1)}) / (t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{whp}{}^{(j)} - P_{whp}{}^{(j-1)}) \\ \sum_{j=1}^{i-1} (P_{whp}{}^{(j)} - P_{whp}{}^{(j-1)}) \log(t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{whp}{}^{(j)} - P_{whp}{}^{(j-1)})(t^{(i)} - t^{(j)}) \\ \sum_{j=1}^{i-1} (P_{whp}{}^{(j)} - P_{whp}{}^{(j-1)}) / (t^{(i)} - t^{(j)}) \\ P_{bhp}{}^{(i)} / P_{whp}{}^{(i)} \\ P_{bhp}{}^{(i)} - P_{whp}{}^{(i)} \\ t^{(i)} \end{bmatrix}, i = 1, \ldots, m \tag{2}$$

The features were derived from pressure response and time data. The feature were then used in linear regression model training together with regularization. The purpose of introducing regularization parameter λ into training was to shrink the weight estimates θ. Hence, the variance of the output can be limited to minimize overfitting issue [13]. Overfitting issue occurs when the model computed is unable to generalize well on new dataset even though the model fits well on the training dataset. Overfitting is usually caused by excessive introduction of features relative to the complexity of the problem.

Linear regression was formulated as Eq.3, and the weight θ was solved using closed-form Eq.4 to minimize the cost function J, as shown in Eq.5 [20]. The advantage of closed-form solution is that it is relatively faster than iterative approach. However, the limitation of the closed-form solution arises when the data is sufficiently larger than the capability of computation memory, or the data form a singular matrix. In the case of this study, closed-form solution was applied as none of the constraints was met. The missing flow rate records can be predicted through matrix multiplication of the trained weights and input at their respective timestep as shown in Eq.6.

$$y^{(i)} = \theta^T x^{(i)} \tag{3}$$

$$\theta = (X^T X + \lambda I)^{-1} X^T y \tag{4}$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^T x^{(i)} - y^{(i)})^2 + \lambda \sum_{n=1}^{m} \theta_n{}^2 \tag{5}$$

$$y^{pred\,(i)} = \theta^T x^{pred\,(i)} \tag{6}$$

The workflow of the study was illustrated in Fig.1.

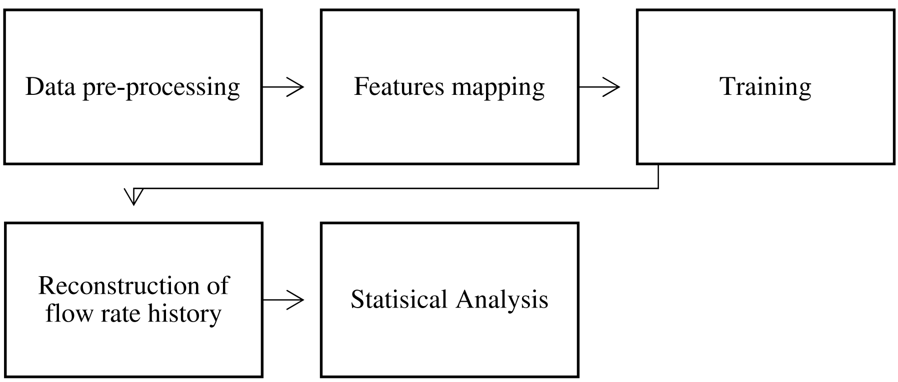

Fig. 1. Workflow of the study

- Data pre-processing: Artificial noise was introduced into synthetic dataset to simulate the slight variation that would have presented in real data. Outliers were removed in the real dataset.
- Features mapping: Extract features based on pressure with respect to time at each timestep. Additional parameters were introduced to model complex dataset as shown in Eq.1 and Eq.2.

- Training: 85% of the data was used as training set to train the model using linear regression.
- Reconstruction of flow rate history: The missing flow rate history will be reconstructed. The remaining 15% of the data that has been segregated from training set will be reconstructed and compared to the actual value to assess the reliability of the linear regression algorithm.
- Statistical Analysis: Statistical analysis such as t-test and p-value evaluation will be carried out to assess the regression model.

5 Results and Discussion

Both synthetic and real dataset were used in this study for flow rate history reconstruction. Linear regression will be first applied on synthetic dataset, then on real dataset. Synthetic dataset had minimal variation, thus suitable to assess the algorithm validity on flow rate history reconstruction. Real dataset had a comparatively higher degree of variation in nature, thus suitable to assess the algorithm's resilience against complexity of the dataset.

5.1 Linear Regression on Synthetic Dataset

Linear regression was first applied on simple synthetic dataset composed of bottomhole pressure, gas flow rate and oil flow rate for a single well with minimal variation in liquid flow rates.

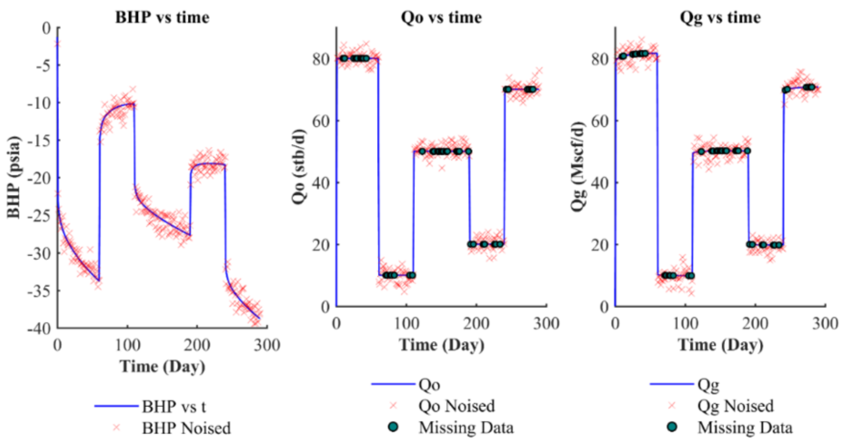

Fig. 2. True and noised pressure and flow rate versus time for synthetic dataset. Missing flow rate history was denoted as green colored dots

Artificial uniformly-distributed noise was added into the synthetic dataset prior to training. The purpose of adding artificial noise into the dataset is to simulate noises that would have been presented in the real dataset collected from the real field. Therefore, we can access the algorithm's performance in making reliable reconstruction of missing data on top of noised dataset.

15% of both the flow rate data, Q_o and Q_g were randomly removed from the synthetic dataset to produce gaps with the absence of flow rate measurements. The missing data on flow rate data were denoted by green dots as shown in Fig.2.

The model was trained with the noised dataset. The trained model showed R^2 value of 0.94 when compared to the original data without noise. The residuals from the training dataset was normally distributed with its mean centered at zero. Thus, it suggested the model could capture most of the points correctly with acceptable variance even though artificial noise was introduced into the system.

As the ground truth of the data is known, cross-validation test was not conducted for model assessment on synthetic dataset. This is because we can directly assess the accuracy by comparing predicted value to the ground truth data. By comparing the reconstructed points with the ground truth data, it was found that the reconstructed flow rate history fits the actual ground truth data with a R^2 value of 0.91. The result demonstrated that linear regression is capable of reconstructing missing flow rate data with low error upon simple synthetic dataset.

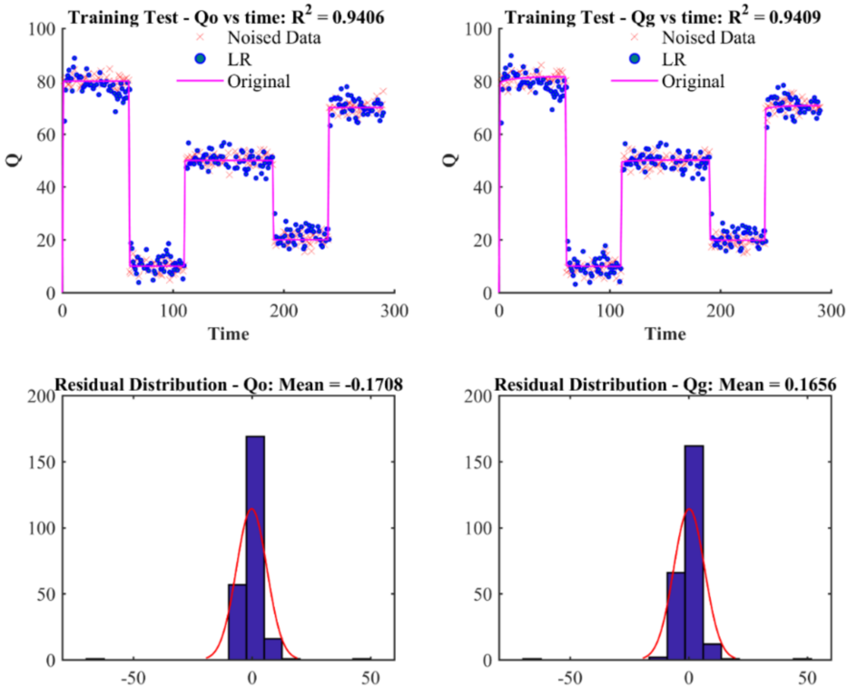

Fig. 3. Comparison between prediction from trained model (blue) with the actual dataset (purple) and residual distribution of trained model for both oil and gas flow rate.

Linear regression is advantageous in term of its interpretability. The statistical tests for each feature provide us a glimpse about the association of each feature to the target value. Parameter estimate, standard error, t-statistic and p-value were the results of the statistical test of the regression. Particularly, the p-value provides crucial insight on how strongly the feature associate with the target value. Features that have p-value lower than 0.01 are statistically significant, which mean that the particular features are meaningful additions to the predictive model.

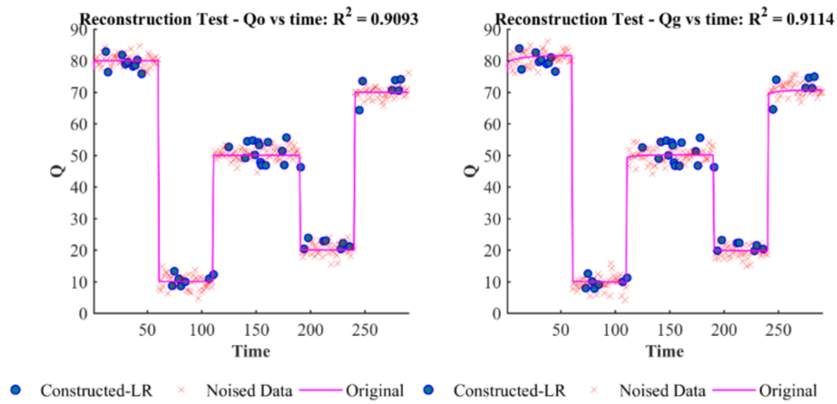

Fig. 4. Comparison between reconstructed flow rate history (blue) with the actual dataset (purple).

From both oil and gas flow rate regression models, the statistical tests showed that the features, P, Plog(t) and P/t had strong association with the regression model, which indicated the presence of superposition events, infinite acting radial flow and the absence of boundary effect upon the reservoir flow [4].

Table 1. Statistical analysis of oil flow rate regression model on synthetic dataset

	Estimate	Standard error	t-statistic	p-value
Intercept	47.476	0.44171	107.48	1.05E-203
P	-34.791	1.4411	-24.143	6.32E-66
Plog(t)	14.46	2.3136	6.2499	1.88E-09
Pt	4.4856	7.8462	0.57169	0.56807
P/t	3.7721	0.97342	3.8751	0.000138
t	3.0384	6.9661	0.43617	0.66311

Table 2. Statistical analysis of gas flow rate regression model on synthetic dataset.

	Estimate	Standard error	t-statistic	p-value
Intercept	47.418	0.44643	1.06E+02	1.66E-202
P	-35.843	1.46E+00	-24.61	2.43E-67
Plog(t)	15.169	2.34E+00	6.4872	5.02E-10
Pt	2.4399	7.9301	0.30768	0.75859
P/t	4.1286	0.98383	4.1965	3.83E-05
t	1.1287	7.0406	0.16032	0.87277

In short, linear regression has an adequate performance on reconstruction of missing flow rate data upon synthetic dataset. However, the pressure and flow rate responses in the real field often fluctuate because of numerous operational reasons and the complex interaction between fluid and the subsurface formation. Hence, the application of linear regression on flow rate history reconstruction should be further investigated using dataset collected from a real producing field.

5.2 Linear Regression on Real Dataset

In the case of real dataset, the pressure response collected over eight years has higher degree of fluctuation compared to that of synthetic dataset.

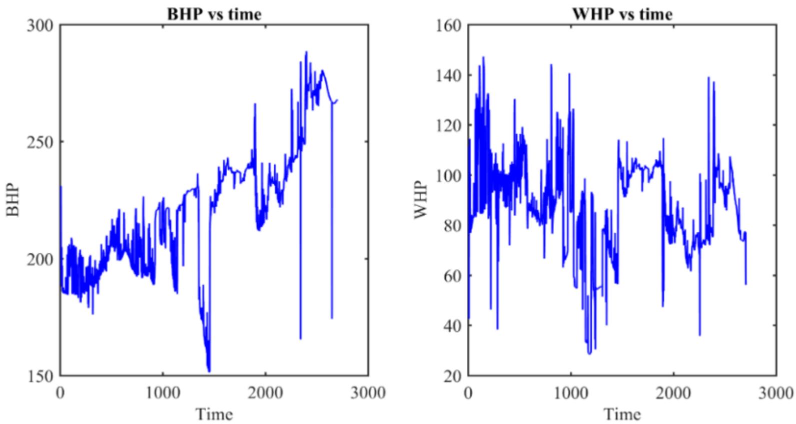

Fig. 5. Bottomhole pressure and wellhead pressure versus time for real dataset

The model was trained with the 90% of the available dataset. The trained model showed R^2 value of 0.41 to 0.63. The residuals from the training dataset was normally distributed with its mean centered at zero, which suggested the model could capture most of the points correctly. The low goodness-of-fit noticed in water flow rate data could be resulted from the inconsistent pressure responses associated with water flow rate in the dataset. The possible solution to this inconsistency could be the inspection of operation history, so that the trend of water flow rate and pressure response can be validated according to the operation conducted.

The flow rate history was reconstructed using the trained model. Cross-validation was conducted by comparing the reconstructed flow rate of the 15% segregated data to the actual value. The reconstructed data fits the actual data with only R^2 values ranged from 0.31 to 0.42.

The linear regression model tended to miss the fluctuation or the spikes that happened along the timeframe. The reconstructed flow rate history was highly fluctuated associated with the measured pressure responses. Therefore, a further post-processing technique was needed to smooth the curve so that a clearer trend can be observed. Exponential weighted moving average technique was used to smooth the flow rate history to get a clearer trend. As a result, the flow rate history was much clearer and agreed to the actual ground truth data. The complete set of reconstructed flow rate history was illustrated in Fig.7.

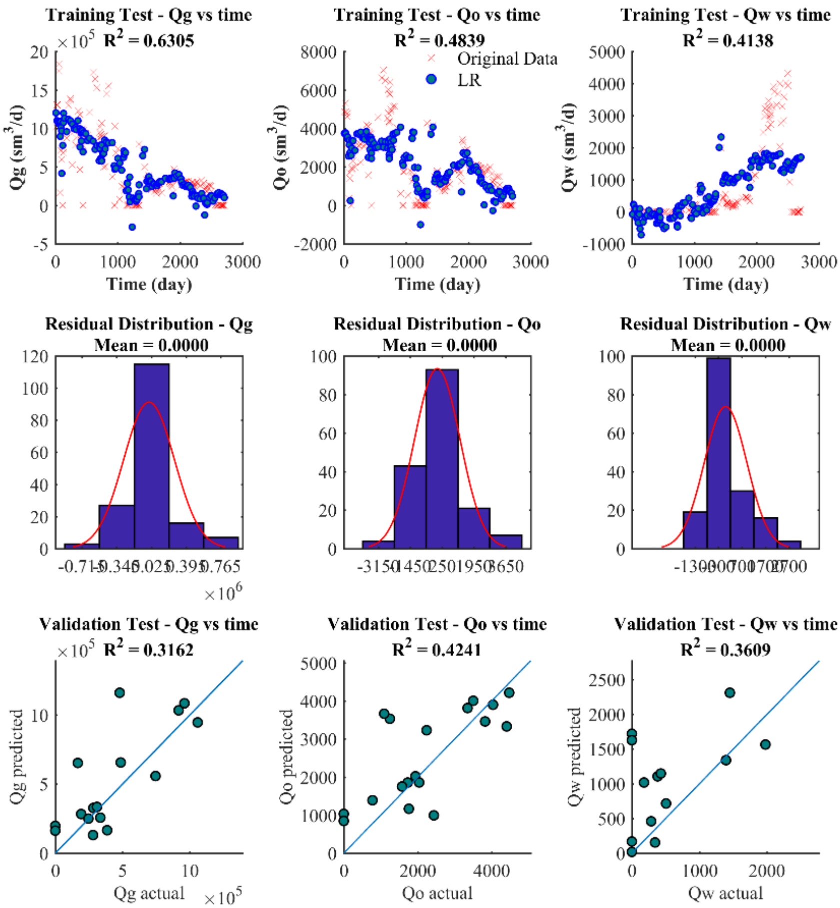

Fig. 6. Comparison between predictions from trained model (blue) with the actual dataset (red) (top). Residual distribution of trained model for both gas, oil and water flow rate (middle). Cross validation of reconstructed flow rate history (green) with the actual data (bottom).

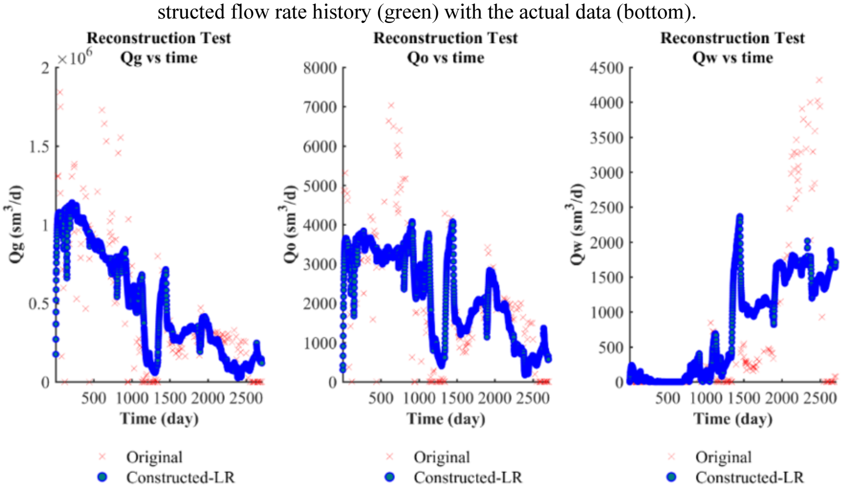

Fig. 7. Comparison between smoothed reconstructed flow rate history (blue) with the actual dataset (red).

The result of the flow rate history reconstruction also showed the limitation of linear regression in modelling highly fluctuated data. The result of flow rate history reconstruction predicted by the model missed out some of the “peaks” that were supposed to present in the system.

Similarly, statistical tests were conducted on oil, gas and water flow rate regression models. Statistical tests of different fluid flows did not show consistent trend as the previous case on synthetic dataset. The observation could be caused by number of reasons, including the existence of complex subsurface reservoir characteristics and the inconsistent pressure response associated with the flow rate due to measurement error. History of production operations will be complementary to the dataset so that the data can be cross-checked to validate the correctness of the data. Nevertheless, a more sophisticated machine learning algorithm is needed to produce better flow rate history reconstruction results on the complex dataset.

Table 3. Statistical analysis of gas flow rate regression model on real dataset

	Estimate	Standard error	t-statistic	p-value
Intercept	4.64E+05	21681	21.379	3.21E-48
P_{bhp}	-77988	1.13E+05	-0.69322	0.48920
P_{bhp} log(t)	1.66E+05	98163	1.6885	0.09332
P_{bhp} t	-9.98E+05	3.82E+05	-2.6113	0.00990
P_{bhp} /t	28433	38873	0.73144	0.46561
P_{whp}	2.20E+05	1.02E+05	2.1553	0.03267
P_{whp} log(t)	-3.22E+05	1.25E+05	-2.5702	0.01110
P_{whp} t	10039	5.76E+05	0.017429	0.98612
P_{whp} /t	-26878	43892	-0.61236	0.54119
P_{bhp} / P_{whp}	-65792	33163	-1.9839	0.04902
$P_{bhp} - P_{whp}$	2.27E+05	99093	2.2944	0.02310
t	9.99E+05	5.83E+05	1.7134	0.088621

Table 4. Statistical analysis of oil flow rate regression model on real dataset

	Estimate	Standard error	t-statistic	p-value
Intercept	2019.2	97.266	20.759	9.63E-47
P_{bhp}	554.02	504.69	1.0977	0.27401
P_{bhp} log(t)	53.919	440.38	0.12244	0.90271
P_{bhp} t	-874.82	1714.6	-0.51022	0.61062
P_{bhp} /t	-156.19	174.39	-0.89565	0.37182
P_{whp}	730.85	457.33	1.5981	0.11205
P_{whp} log(t)	-941.8	561.63	-1.6769	0.095565
P_{whp} t	-5203.5	2584	-2.0137	0.045757
P_{whp} /t	-51.237	196.91	-0.26021	0.79504
P_{bhp} / P_{whp}	-339.02	148.78	-2.2787	0.024039
$P_{bhp} - P_{whp}$	1086.9	444.55	2.445	0.015597
t	6848.1	2615.7	2.6181	0.009713

Table 5. Statistical analysis of water flow rate regression model on real dataset

	Estimate	Standard error	t-statistic	p-value
Intercept	689.26	72.488	9.5085	3.40E-17
P_{bhp}	466.37	376.13	1.2399	0.21686
P_{bhp} log(t)	-73.809	328.2	-0.2249	0.82236
P_{bhp} t	999.79	1277.8	0.7824	0.43515
P_{bhp} /t	-19.008	129.96	-0.1463	0.88391
P_{whp}	350.66	340.83	1.0288	0.30516
P_{whp} log(t)	-428.57	418.56	-1.0239	0.30746
P_{whp} t	3833.8	1925.8	1.9908	0.04825
P_{whp} /t	-188.93	146.75	-1.2875	0.19984
P_{bhp} / P_{whp}	30.763	110.88	0.2775	0.78180
$P_{bhp} - P_{whp}$	180.6	331.3	0.5451	0.58644
t	-3378.9	1949.4	-1.7333	0.08501

6 Conclusion

In this study, we aimed to investigate the efficacy of linear regression in flow rate history reconstruction. The use of linear regression allowed us to extract and expand features from pressure-flow rate data based on physical laws governing fluid flow in porous media. Our results indicate that linear regression is a suitable approach for modeling datasets with minimal variation, achieving an R^2 value of up to 0.91. However, we found that linear regression struggled to model datasets with a high degree of variation, as evidenced by the real dataset, which exhibited a significant drop in goodness of fit to an R^2 value of 0.37. This suggests that while linear regression can be effective for simpler pressure-flow rate datasets, it may be inadequate for more complex datasets that exhibit significant fluctuations in the flow rate. Despite these limitations, the use of linear regression for flow rate history reconstruction remains valuable due to its fast computation time and high interpretability. In summary, this study highlights the potential of linear regression as a tool for flow rate history reconstruction, particularly for simpler datasets. However, researchers must carefully consider the limitations of this approach and explore alternative techniques to improve the accuracy and reliability of flow rate history reconstruction on more complex datasets.

7 Future Research Directions

A future direction for research in flow rate history reconstruction is to investigate the development of a feature set that can capture a more meaningful representation of pressure-flow rate data. Additionally, the application of alternative machine learning techniques, such as neural network modeling, is recommended to potentially yield a more robust predictive model for flow rate history reconstruction. The utilization of these techniques may offer increased resilience against fluctuations in the pressure-flow rate dataset, ultimately improving the accuracy and reliability of flow rate history reconstruction.

Nomenclature

P	pressure
P_{bhp}	bottomhole pressure
P_{whp}	wellhead pressure
t	time
Q	flow rate
x	features for one observation.
x^{pred}	features of one observation for prediction
X	feature matrix
y	output for one observation
y^{pred}	output of one observation for prediction
θ	weights for features x
λ	regularization parameter
m	number of observations
n	number of features

Acknowledgement

We would like to express our gratitude towards Petroleum Engineering Department and Institute of Hydrocarbon Recovery for the support given throughout this research.

References

1. Junqueira CP, Ortiz CEP (2013) Reservoir parameters estimation and flow rate history reconstruction using downhole pressure data. Paper presented at the *22nd International Congress of Mechanical Engineering (COBEM 2013)*, Ribeirão Preto, SP, Brazil, 3-7 November 2013
2. Athichanagorn S, Horne RN, Kikani J (1999) Processing and interpretation of long-term data from permanent downhole pressure gauges. Paper presented at the *SPE Annual Technical Conference and Exhibition*, Houston, Texas, 3-6 October 1999
3. Martin E, Wills P, Hohl D, Lopez JL (2017) Using machine learning to predict production at a Peace River Thermal EOR Site. Paper presented at the *SPE Reservoir Simulation Conference*, Montgomery, Texas, USA, 20-22 February 2017
4. Cao Q, Banerjee R, Gupta S, Li J, Zhou W, Jeyachandra B (2016) Data driven production forecasting using machine learning. Paper presented at the *SPE Argentina Exploration and Production of Unconventional Resources Symposium*, Buenos Aires, Argentina, 1-3 June 2016
5. Li Y, Han Y (2017) Decline curve analysis for production forecasting based on machine learning. Paper presented at the *SPE Symposium: Production Enhancement and Cost Optimization*, Kuala Lumpur, Malaysia, 7-8 November 2017
6. Subrahmanya N, Xu Y, El-Bakry A, Reynolds C (2014) Advanced machine learning methods for production data pattern recognition. Paper presented at the *SPE Intelligent Energy Conference & Exhibition*, Utrecht, The Netherlands, 1-3 April 2014
7. Ramgulam A, Ertekin T, Flemings PB (2007) An artificial neural network utility for the optimization of history matching process. Paper presented at the *Latin American & Caribbean Petroleum Engineering Conference*, Buenos Aires, Argentina, 15-18 April 2007

8. Aulia A, Jeong D, Saaid IM, Shuker MT, El-Khatib NA (2017) A new history matching sensitivity analysis framework with random forests and Plackett-Burman Design. Paper presented at the *SPE Symposium: Production Enhancement and Cost Optimisation*, Kuala Lumpur, Malaysia, 7-8 November 2017
9. Tian C, Horne RN (2015) Machine learning applied to multiwell test analysis and flow rate reconstruction. Paper presented at *SPE Annual Technical Conference and Exhibition*, Houston, Texas, USA, 28-30 September 2015
10. Gaganis V, Varotsis N (2012) Machine learning methods to speed up compositional reservoir simulation. Paper presented at the *SPE Europec/EAGE Annual QConference*, Copenhagen, Denmark, 4-7 June 2012
11. Tian C, Horne RN (2015) Applying machine learning techniques to interpret flow rate, Pressure and temperature data from permanent downhole gauges. Paper presented at the *SPE Western Regional Meeting*, Garden Grove, California, USA, 27-30 April 2015
12. Horne RN (2007) Listening to the reservoir—interpreting data from permanent downhole gauges. *Journal of Petroleum Technology,* 59(12): 78-86
13. Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning*. Springer series in statistics, New York
14. Horne RN, Liu Y (2013) Interpreting pressure and flow rate data from permanent downhole gauges using convolution-kernel-based data mining approaches. Paper presented at the *SPE Western Regional & AAPG Pacific Section Meeting 2013 Joint Technical Conference*, Denver, Colorado, USA, 30 October-2 November 2013
15. P Almeida, Carrasquilla A (2016) Integrating geological attributes with a multiple linear regression of geophysical well logs to estimate the permeability of carbonate reservoirs in Campos Basin, Southeastern Brazil. *Analyst,* 9(4): 8-17
16. Peng Z, Yang H, Pan H, Ji Y (2016) Identification of low resistivity oil and gas reservoirs with multiple linear regression model. Paper presented at the *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD),* Changsha, China, 13-15 August 2016
17. Asumadu-Sarkodie S, Owusu PA (2017) Recent evidence of the relationship between carbon dioxide emissions, energy use, GDP, and population in Ghana: A linear regression approach. *Energy Sources, Part B: Economics, Planning, and Policy,* 12(6):495-503
18. Clement D, Gruber N (2017) AC*-based extended multiple linear regression method to determine decadal changes in anthropogenic CO_2 in the ocean. In: *19th EGU General Assembly Conference Abstracts*, Vienna, Austria, 23-28 April 2017, vol 19, p 17207
19. Pino-Mejías R, Pérez-Fargallo A, Rubio-Bellido C, Pulido-Arcas JA (2017) Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO_2 emissions. *Energy,* 118: 24-36
20. Hastie T, Tibshirani R, Friedman J (2009) Overview of supervised learning. In: Hastie T (ed) *The elements of statistical learning*. Springer, p 9-41