Prediction of Two-phase Flow-rate through Wellhead Chokes in Oil Wells Using Machine Learning

Negar Nazari (nazari@Stanford.edu), Talal Al Shafloot (talalss@Stanford.edu)

DEPARTMENT OF ENERGY RESOURCES ENGINEERING, STANFORD UNIVERSITY

Introduction

Introduction:

- Wellhead assembly is an essential part of a producing oil or gas well that protects downstream facilities from the danger of high flow rates.
- Choke controls the flow rate of multiphase flux, and protects the hydrocarbon formation and surface equipment from probable fluctuation in pressure.

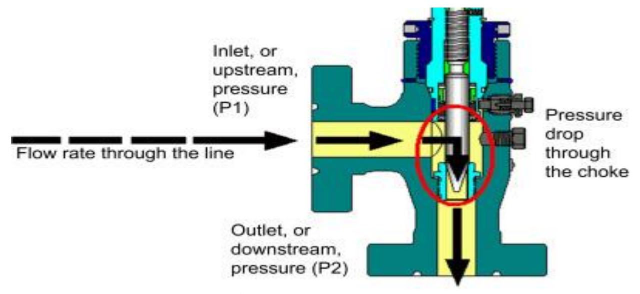

Fig. 1: Schematic of choke https://www.sciencedirect.com/topics/earth-and-planetary-sciences/choke (Courtesy from Cameron)

- Accurate prediction of flow rate through chokes is extremely helpful for assessing the reservoir performance and production forecasting and it is essential for establishing a controllable and stable flow in producing wells.
- Flow-meters are expensive and hard to be implemented for large fields, hence, measuring the production rate of an oil well is hard. Furthermore, in fields with advanced well systems, multiple wells are connected to one manifold, and the flow rate reported from the manifold is for all wells and not for a single one.
- Due to the simplicity of the current correlations and their sensitivity to choke size, they are not able to predict the flow accurately.
- As a result, a robust model is needed to estimate flow rate in oil wells from simple parameters measured at the wellhead assembly.

Background:

- Gilbert presented the most common correlation to calculate flow rates through chokes as:

$$Q_l = \frac{A \times P_{up}^{D} \times S^{B}}{GLR^{C}}$$

where Q_l is liquid gross flow rate, P_{up} is the upstream pressure, GLR is the gas liquid ratio, and S is the choke size as a multiple of (1/64) inches. A, B, C and D are fitting parameters.

- Recent studies are conducted on small set of datasets and none of them included complex models. Furthermore, no studies were done to generate a model applicable on different fields with various flow and formation properties.

Data Processing

Features:

Old Models: upstream pressure (P_{up}), gas liquid ratio (GLR), choke size (S)

Our Models: Old Features **+** Differential pressure ($\Delta P = P_{up} - P_{dn}$), gas Oil ratio ($GOR$), Temperature($T$), Water-cut ($WC$), Flow regime ($Critical$ or $Subcritical$)

Data Filtering: Data beyond 3σ are removed.

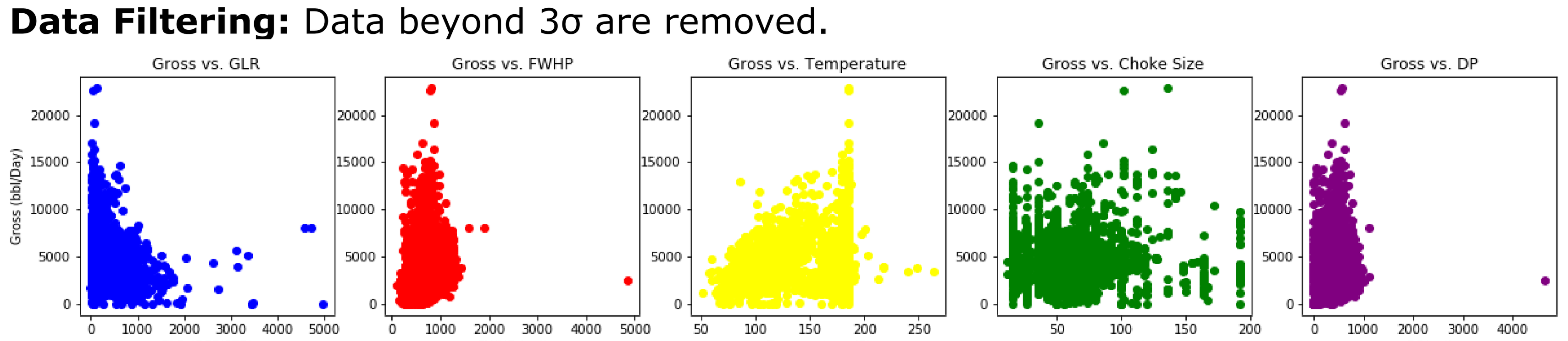

Fig. 2: Gross vs. different features before filtering

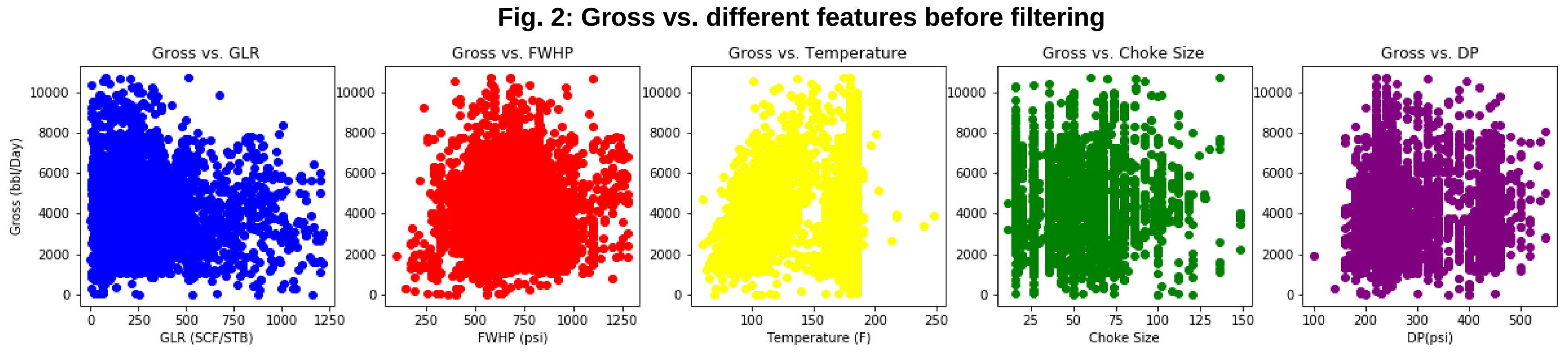

Fig. 3: Gross vs. different features after filtering

Notes on Choosing Features:

- **Darcy's Law:**
  i. **q ∝ P => Pressure was removed from the features and it's exponent was fixed to 1.**

$$q = \frac{kA}{\mu}\frac{\Delta P}{L}$$

- **Flow Regime:**
  i. **The P_{dn}/P_{up} ratio determines the flow regime to be Critical (Ratio<=0.5) or Subcritical ((Ratio>0.5)) and it is introduced as a feature.**

Models

Linear Models:

- **Linear/Ridge Regression:** A baseline to fit a linear model with coefficients to minimize L, w/wo the L_2−norm regularization, respectively.

$$L = ||X\omega - \hat{y}||_2^2 \quad \& \quad L = ||X\omega - \hat{y}||_2^2 + \alpha||\omega||_2^2$$

- **Bayesian Ridge Regression:** Uses probability distributions rather than point estimates. The prior for the coefficient is given by a spherical Gaussian as:

$$p(\omega|\lambda) = N(\omega|0, \lambda^{-1}I_p)$$

- **Polynomial Linear/Ridge Reg. : F**ormulates the model using an n^{th} degree polynomial to minimize L, w/wo the L_2−norm regularization , respectively.

Neural Network Model:

- **Multi-Layer Perceptron:** The activation function for hidden layers is ReLU function. It uses backpropagation and the loss function is defined as:

$$L = \frac{1}{2}||y - \hat{y}||_2^2 + \frac{\alpha}{2}||\omega||_2^2$$

Nearest Neighbor Model:

- **K-Nearest Neighbor Regression:** The target is predicted by local interpolation of the targets associated of the nearest neighbors in the training set.

Ensemble Methods Models:

- **Random Forrest Regression:** Makes predictions by combining decisions from a sequence of base models as:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \cdots$$

- **Gradient Tree Boosting:** Builds an additive forward stage-wise model to allow the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.
- **Extra Tree Regression:** Implements a meta- estimator to fit a number of randomized decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

$$\hat{F}(x) = \arg\min_{y} E_{x,y}[L(y, F(x))]$$

Results

Steps:

- We have used 10 different models from various families for this study. Total of dataset size was 4677, and reduced to 4063 after excluding the data beyond 3σ.
- After random permutation, 80% of the data were used for training and 20% for testing.
- The dataset was preprocessed by transforming the data to center by removing the mean value of each feature, and then scaling it by dividing non-constant features by their standard deviation.
- The hyperparameter tuning was implemented for the best model to optimize its performance and improving the results. The data was divided to 80, 10, 10 for training, testing, and validation, respectively.

All Models on All Three Fields:

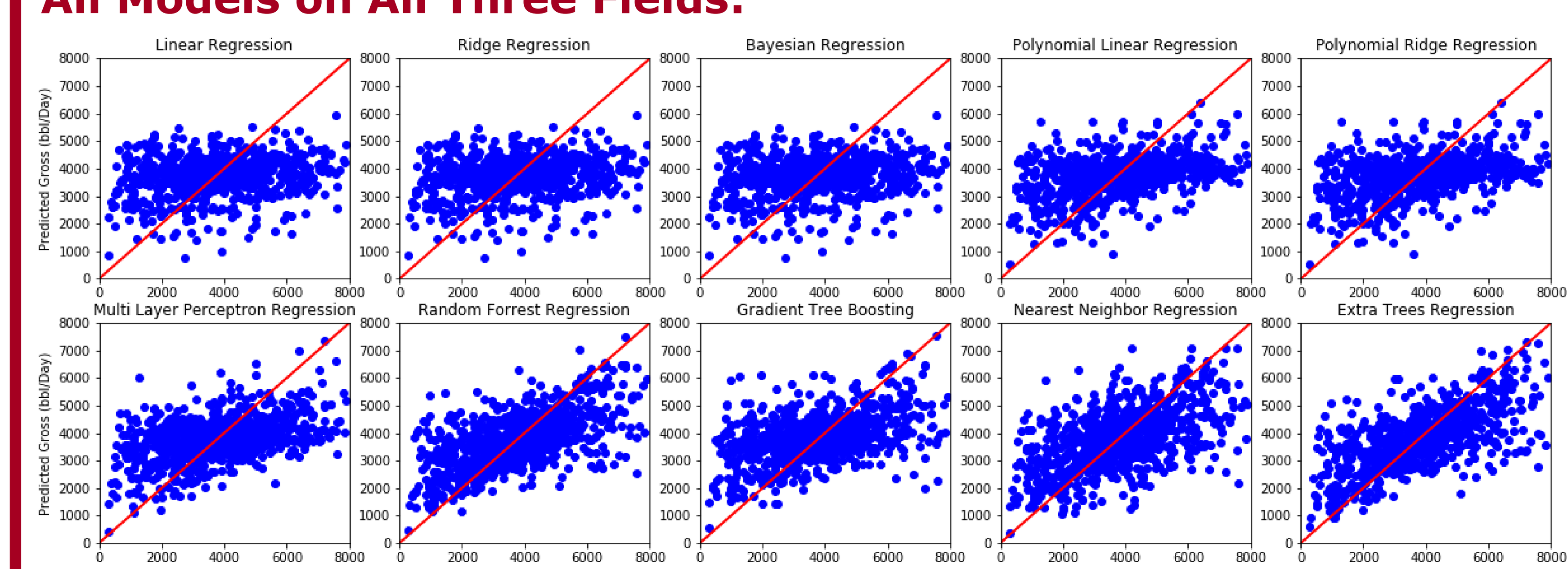

Fig. 4: Predicted vs. measured gross for all three fields obtained from all models.

Results

Top-3 Models Result for All Three Fields:

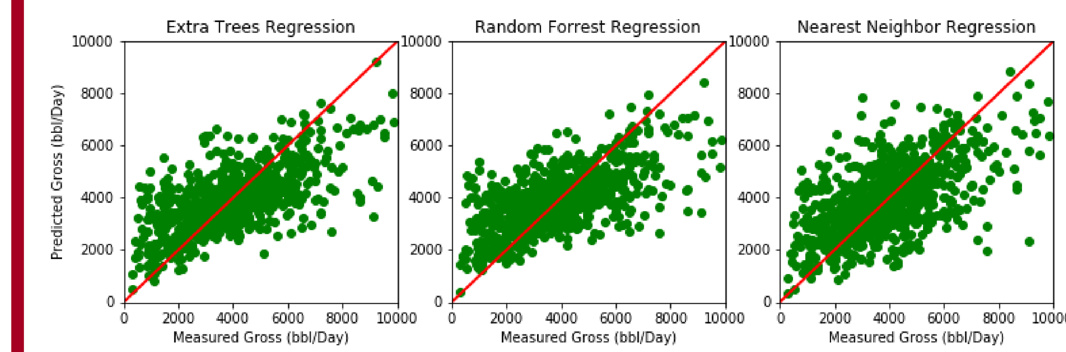

Fig. 5: Predicted vs. measured gross for all three fields obtained from 3 top models.

Table 1: Statistics of top 3 models for all three fields.

Top-3 Model Performances for all 3 Fields			
Model	Train Score	Test Score	Correlation Coefficient
Extra Tree Regressor	0.99999	0.55322	0.67435
Random Forest Regressor	0.94050	0.53192	0.64741
Nearest Neighbor Regressor	0.76366	0.45276	0.60552

Top-3 Models Result for Fields A & C:

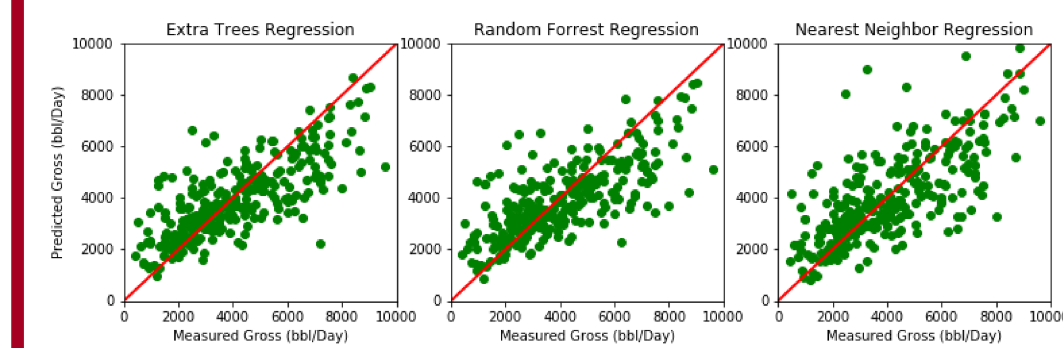

Fig. 6: Predicted vs. measured gross for fields A & C obtained from 3 top models.

Table 2: Statistics of top 3 models for fields A & C.

Top-3 Model Performances for Fields A and C			
Model	Train Score	Test Score	Correlation Coefficient
Extra Tree Regressor	1.00000	0.71040	0.75842
Random Forest Regressor	0.94690	0.68243	0.73071
Nearest Neighbor Regressor	0.77634	0.63726	0.71171

Top-3 Models Result for Field B:

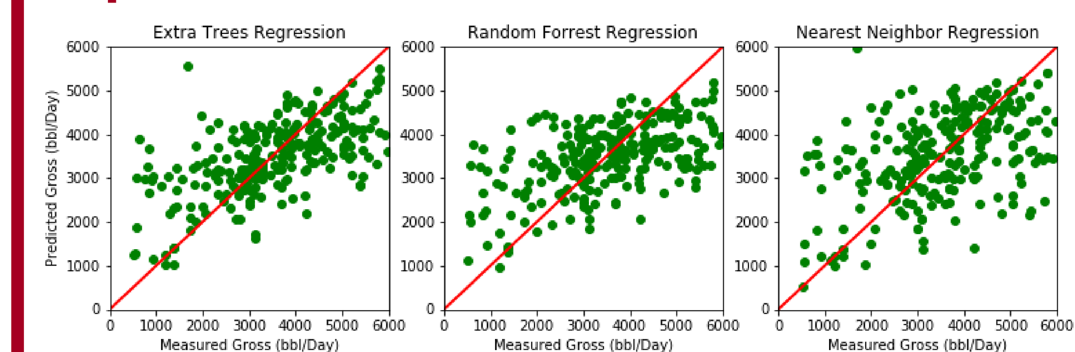

Fig. 7: Predicted vs. measured gross for field B obtained from 3 top models.

Table 3: Statistics of top 3 models for field B.

Top-3 Model Performances for Field B			
Model	Train Score	Test Score	Correlation Coefficient
Extra Tree Regressor	1.00000	0.48347	0.56972
Random Forest Regressor	0.90986	0.33624	0.43820
Nearest Neighbor Regressor	0.70280	0.30474	0.44213

Discussion, Conclusion and Future Work

Discussion:

- Linear models are not capable of capturing the nonlinear behaviors, hence, they show a weak performance (R2-score <=0.2).
- The best performance in the linear models family belongs to polynomial models, as higher degree polynomials are capable of capturing different behaviors. However, due to the overfitting issue, their performance improvement is extremely limited.
- Applying the neural network models improves the R2-score up to 40% for some cases, but more complexed models need to be developed.
- The nearest neighbor and ensemble models show the most optimized performance for all cases. Tuning the hyperparameters results in capturing complex behaviors.
- Dataset B shows a poor results, which implies that more features related to flow and formation properties must be added to capture all different aspects of the behavior.

Conclusion:

- Applying Gillbert correlation to predict gives an score of 0.13, which is less than all the models that we developed, considering that this was the most complicated dataset that has been used for flow prediction through choke.
- The neural network models can be improved by adding more hidden layers with different activation functions.
- Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model.
- This work can be applied for water resources study and reduces the costs of flow behavior significantly.

Future Work:

- Due to the strange behavior of field B, we will run CFD to validate this dataset.
- A new neural network model with more complex hidden layers will be implemented
- Further tuning of hyperparameters may helps improving the performance of the ensemble models and it will be discovered later to generate the most optimized predictive model for oil and gas production.
- For the best predictive mode, flow and formation properties will be added.