Recent studies are conducted on small set of datasets and none of them included Bayesian Gradient Gilbert presented and may helps improving the performance of the Extra Tree Regression: Flow The further tuning of the neural network models can be improved by adding more hidden layers with Results Ensemble learning helps improve machine learning results by combining several, respectively.

Darcy’s Law:

\[ \frac{Q}{L} = \frac{k}{\mu} \frac{\Delta P}{L} \]

where \( Q \) is the linear flow rate, \( k \) is the permeability, \( \mu \) is the fluid viscosity, \( \Delta P \) is the pressure gradient, and \( L \) is the distance.

OF STANFORD P => Pressure was removed from the features is liquid gross

Choke controls Due to the strange behavior of field B, we will run CFD to validate this dataset.

Old Features Water

New Features Flow

Fig. 2: Gross vs. different features before filtering

Fig. 3: Gross vs. different features after filtering

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

<table>
<thead>
<tr>
<th>Field</th>
<th>Model</th>
<th>MSE</th>
<th>Feature Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Fitted GPR</td>
<td>0.012</td>
<td>0.43, 0.47, 0.31</td>
</tr>
<tr>
<td>B</td>
<td>Fitted GPR</td>
<td>0.015</td>
<td>0.39, 0.42, 0.25</td>
</tr>
<tr>
<td>C</td>
<td>Fitted GPR</td>
<td>0.014</td>
<td>0.38, 0.40, 0.26</td>
</tr>
</tbody>
</table>

All Models on All Three Fields:

<table>
<thead>
<tr>
<th>Field</th>
<th>Model</th>
<th>MSE</th>
<th>Feature Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Fitted GPR</td>
<td>0.012</td>
<td>0.43, 0.47, 0.31</td>
</tr>
<tr>
<td>B</td>
<td>Fitted GPR</td>
<td>0.015</td>
<td>0.39, 0.42, 0.25</td>
</tr>
<tr>
<td>C</td>
<td>Fitted GPR</td>
<td>0.014</td>
<td>0.38, 0.40, 0.26</td>
</tr>
</tbody>
</table>

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 their degree forming polynomial 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 complicated 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 Gilbert 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 help 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.