1) Introduction

2) Geological Overview
3) ML Process (ensuring usability, stability & interpretability)
4) Data Used
5) Usability & Stability
6) Interpretation Tools
7) Conclusions
Spirit River Activity

- Approximately 2600 Horizontal wells have been drilled since 2010.
- 2.6 Bcf/day Production
- Even with AECO prices declining over the years, drilling is still active (Still Economic)
Top 12 Spirit River Producers

Spirit River Cumulative Production by Operator

Cumulative Volume (bcf)

Data provided by HHS Market Information Hub and/or gDCI/CDI - Feb 11, 2019, 3:08 PM VERIZON™
2) Geological Overview

1) Introduction

2) Geological Overview

3) ML Process (ensuring usability, stability & interpretability)

4) Data Used

5) Usability & Stability

6) Interpretation Tools

7) Conclusions
GLJ has evaluated significant portions of the Spirit River across all the development areas. (Grey Lands)
There is no formally agreed on further division of the Notikewin, Falher and Wilrich members, which can cause considerable confusion within industry.

GLJ has adopted a consistent stratigraphic nomenclature across the entire Deep Basin.
Deposition of the Machine learning targets (Wilrich and Falher) occurred in shorefaces which prograded seaward from the south to the Northwest. These shorefaces were also incised with valley fill deposits of the Falher
Over 5000 wells were picked for formation tops.

Core work is the foundation of GLJ’s Petrophysical work.

- 63 publicly available core.
- 14 Core with special analysis.
  - Cap pressure, Salinity, Electrical Properties, XRD, etc.
- Used both with porosity and water saturation (oil based cores) to tie our wireline petrophysics values.

1000s of petrophysical evaluations were completed across the trend.
Subsurface parameters utilized in GLJ’s regional work for each well include:

- Zone ID
- Gross Thickness
- Net Pay
- Net to Gross
- Average Porosity
- Average VShale
- Average Sw
- HCPV
- Depositional Facies
- Petrophysical Sensitivities
- Pressure Gradient
- Temperature Gradient
- iC4/nC4 Ratios
- Condensate Yields
3) Machine Learning Process

(ensuring usability, stability & interpretability)

1) Data Used
2) Usability & Stability
3) Interpretation Tools
4) Conclusions
1) **Target** = what we want to predict (the “right answers” have been provided for Training)

2) **Features** = the inputs we want to use to predict the Target

3) **Training** = the process that uses Features and Targets to create a predictive Model

4) **Model** = the algorithm(s) used to generate Target predictions

5) **Feature Importance** = a measure of how impactful a Feature is on the predictive capability of a Model

6) **$R^2$** = the coefficient of determination (i.e. represents the percent of variance that can be explained by a Model for a set of features/inputs)
Machine Learning Modeling Objectives

1) Predict production performance (Target = 12 month cumulative gas)
   - Understand Feature correlations & dependencies
   - Build an understanding of what matters (measure each Feature’s impact)
   - Measure the predictive limits of the data

2) Identify “roll-over” points and avoid over-capitalization

3) Use the predictive model to test hypotheses

4) Incorporate costs to value-optimize completions for specific reservoir characteristics
Collaboration of technical & domain experts to generate & review dozens of model iterations to:

- Maximize predictive capability
- Select “optimal” Feature set
- Maximize usability & build trust
- Ensure interpretability & buy-in
4) Data Used

1) Introduction
2) Geological Overview
3) ML Process (ensuring usability, stability & interpretability)

4) Data Used

5) Usability & Stability
6) Interpretation Tools
7) Conclusions
Data Coverage

906 wells used in ML model (target & all features populated)
Data generated for this Machine Learning Project

1) **Target information** (condensed 12 month cumulative gas)

2) **Well location information** (e.g. X, Y, TVD, Region, azimuth)

3) **Offset drainage** (production within 400, 800, 1600m radii)

4) **Subsurface information**
   a) **Geological information**: (e.g. Zone, Depositional Facies, map parameters: Vsh, Net Pay, Gross Pay, NtG, porosity, Sw, HCPV, at different Vsh/porosity cutoffs – 40%+3%, 50%+3%, 50%+2%)
   b) **Reservoir fluid information**: (e.g. Pressure, %over/under pressure, temperature + temperature gradient, iC4/nC4 ratio, shallow- and deep-cut condensate yields)

5) **Completion information**
   e.g. Frac technology, base fluid, energizer, balls recovered, estimated problem time, fluid volume+concentration, %fluid recovery, #stages, spacing, proppant sizes, tonnage placed by proppant type, total proppant intensity (t/m), proppant intensity by proppant type, completed hz length
5) Usability & Stability

1) Introduction
2) Geological Overview
3) ML Process (ensuring usability, stability & interpretability)
4) Data Used

5) Usability & Stability

6) Interpretation Tools
7) Conclusions
What’s an ideal Feature count?

**51 Feature Model**

FirstTwelveMonthsGasMMcf

$R^2 = 0.58$

**13 Feature Model**

FirstTwelveMonthsGasMMcf

$R^2 = 0.53$

**6 Feature Model**

FirstTwelveMonthsGasMMcf

$R^2 = 0.47$
What’s an ideal Feature count?

Recursive Feature Elimination using a simple model

Score = r^2

Number of Features Selected

6 13 51

It’s not just about how many features are used, but which features are selected.
Feature Selection Goals

- Maximize **stability** (minimize data redundancy, avoid highly correlated features)

- Maximize **usability** (without compromising predictive capability)

- Support **interpretability**

- Ensure **buy-in**
6) Interpretation Tools

1) Introduction
2) Geological Overview
3) ML Process (ensuring usability, stability & interpretability)
4) Data Used
5) Usability & Stability

6) Interpretation Tools

7) Conclusions
### Before you build a model: Feature Correlation Matrix

- **Used to identify correlations (data redundancy) between Features**
- **Helps inform the Feature selection process**

```
<table>
<thead>
<tr>
<th>Variable 1</th>
<th>Variable 2</th>
<th>Variable 3</th>
<th>Variable 4</th>
<th>Variable 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable A</td>
<td>-0.56</td>
<td>0.13</td>
<td>-0.08</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>-0.01</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>-0.13</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>-0.36</td>
<td>0.12</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>-0.05</td>
<td>0.10</td>
<td>-0.20</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>-0.15</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>-0.09</td>
<td>0.12</td>
<td>-0.03</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>-0.05</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>-0.20</td>
<td>0.11</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

```
We have a trained model... now what?

- Target = 12 month cumulative gas (MMcf)
- 13 Feature Model
- out-of-sample $R^2 = 0.53$

Now we can put the model through its paces to generate predictions and build interpretive tools.
Feature Grouping can help us understand which broader factors matter most (e.g., geology, pressure, lateral length, completion design parameters)

- Helps answer questions: e.g., how important are all the completion features in aggregate versus all the geological features in aggregate?

- Treating features as a group removes the effects of data leakage between features within that group
Feature Importance Using Grouping

Feature Grouping helps to craft a meaningful story
Completion = Controllable

Subsurface = Non-controllable (but selectable... you can choose where you drill your well)
ICE & IFLE

How to use Machine Learning to build understanding & fuel explanations
ICE Plot: Context and Goal

**Context:**
1) We have trained a Machine Learning Model to predict a **Target** using several **Features**
2) **Target** for this example = **Production Performance** (e.g. 12 month cum)
3) **Feature** for this example = **Proppant Intensity** (t/m)

**Goal:**
1) Characterize the Production Performance (Target) **response to changes** in Proppant Intensity (Feature) over the range of Proppant Intensity values in the dataset.
2) Leverage these visual tools to build trust, understanding and actionable insight.
1) Predict the Production for a well using its actual input values (Features)

2) Next step → “turn the dial” for that well, on only one Feature (i.e. Proppant Intensity), keeping all other input values the same, and generate predictions for the range of Proppant Intensity values in the dataset…
1) The resulting blue line shows how predictions for one well’s Production changes as Proppant Intensity is varied. This is called an “ICE Plot”.

2) Next step → apply this to all wells…
1) Generate ICE lines for all wells to see patterns of Production response to changes in Proppant Intensity

2) Next step → superimpose an average prediction for all wells…
**Individual Conditional Expectation (ICE) Plot**

- **Blue lines**: individual well predictions
- **Red X**: each well’s actual Proppant value
- **Yellow line**: average Production prediction of all wells for each quantile of the Proppant values
- **Black dots**: Proppant quantiles used to generate the predictions
- **Dashed-red line**: average Production value
- **Maximum Average Impact** caused by changing Proppant values over the range of values in the dataset
- **Distribution of Data Points** for Proppant (Feature of interest)

**Distribution of Data Points used to bin into 50 quantiles**
ICE Plot Average Line = Partial Dependence Plot (PDP)

ICE plots show individual well impacts, patterns of response & divergence in individual well response.
Response to Proppant Intensity varies by geologic zone

**Uncentered PDP**  
(impact in the context of actual values)

**Centered PDP**  
(relative impact, amplifies the shape)
**Linear Model vs Machine Learning Model**

**Linear model**  
(multi-linear regression)  
$R^2 = 0.36$

- **Limited interpretability**

**Machine Learning model**  
$R^2 = 0.53$

- **High interpretability**

Suggests roll-over point
ICE Plots help you identify impact (i.e. what matters)

**Big Impact**

**Small Impact**
Introduction to IFLE

ICE (Individual Conditional Expectation)
- shows the production response to changes in an individual feature over the range of values in the dataset

IFLE (Individual Feature Localized Effect)
- feature attribution method (based on game theory)
- quantifies the production effect that each feature value has on each well (i.e. it explains each well’s production performance feature by feature)
- that effect is measured in MMcf relative to the average well
Individual Feature Localized Effects (IFLE)

- Pay effect is negative.
- Hz length effect is negative.
- Aggregate effect is 6% below average well.
- Proppant value is adding to production performance.

You’re in a good zone.
Individual Feature Localized Effect (IFLE) of Proppant Intensity

- Calculation of the effect **each feature** has on production performance for **each well**
- Measured relative to the average predicted production value
- Quantified explanation of what is driving each well’s production performance (measured in MMcf)

Red bar = Proppant Intensity IFLE value
Grey bars = IFLE value for all other features
Black bar = the well’s aggregate effect (above or below the average well)
Introduce Economics for Completions (Profit) Optimizer

This well could have achieved $1.1 million more in 12 month profit by using 0.8 t/m
Completions Profit Optimization (for Proppant Intensity)

Distribution of well’s lost-profit opportunity

Benchmarking Operators
(percent of wells in each opportunity category)

12-month revenue for optimal proppant intensity, minus 12 month revenue for actual proppant intensity used
7) Conclusions

1) Introduction
2) Geological Overview
3) ML Process (ensuring usability, stability & interpretability)
4) Data Used
5) Usability & Stability
6) Interpretation Tools

7) Conclusions
Conclusions

- **Subsurface (geology & reservoir) matters!**
  - and we can quantify how much it matters
  - In the Spirit River completions & geoscience contribute equally to production prediction

- Consistent geological interpretation is required on a regional basis for effective machine learning

- Integrating geology & engineering teams with machine learning is critical to generating meaningful, usable, interpretable results

- Visualizations are the foundation of interpretability & building trust

- Machine learning offers powerful insights by quantifying (i.e. explaining) what contributes to production performance & informing optimal completion design decisions
Acknowledgments

**GLJ Engineering Support:**
- Scott Quinell – Senior Reservoir Engineer, GLJ
- Bill Spackman – Manager, Engineering, GLJ

**Verdazo Analytics Machine Learning support:**
- Brian Emmerson – Director of Data Science, Verdazo Analytics
- Anton Biryukov – Data Scientist, Verdazo Analytics
- Tyler Schlosser – Senior technical Advisor, Verdazo Analytics
If you would like to learn more about purchasing the detailed machine learning study and predictive model contact sales@verdazo.com
Appendix

supportive information
GLJ Petroleum Consultants is a Calgary-based oil & gas consulting firm providing independent petroleum reserves evaluation & energy consulting services. With over 45 years’ experience in Canada, and across the Globe, GLJ has become one of the largest reserve evaluation firms in the world. While working alongside clients to help elevate business decisions, attract and optimize capital, GLJ leverages our expertise and rich geological and engineering resource play data sets in plays such as the Spirit River, Bakken, Cardium, Duvernay, Montney, Oil Sands and Viking.

Verdazo Analytics is a Calgary-based software and consulting company focused exclusively on Oil & Gas related visual analytics & machine learning. Verdazo covers all aspects of the asset life cycle including planning, drilling, completions and operations. Verdazo has helped more than 130 companies, over 12 years, to use analytics to maximize value. Recently acquired by Pason, Verdazo is growing its offering in machine learning and its presence in the US.
Spatial Residual Analysis (906 wells)

**Abs Percent Error**

- **Absolute % Error <100** = 831 wells
- **Absolute % Error <40%** = 620 wells (shown in red)
- **Absolute % Error >100%** = 75 wells (shown in red)

831 wells (quartile colouring)
Impact Importance Plot: Summary of ICE Plot Impact for all Features

Low (30% below avg)  Avg Production  High (38% above avg)

- HZLength(m)
- TotalProppantIntensity(\_m)
- Zone_LABELS
- ZoneTVD(m)
- PercentOver-UnderPressure
- 40PercentVsh3PercentPhi-V...
- 40PercentVsh3PercentPhi-NG
- FracSpacing(m)
- ProppantConcentration(\_m3)
- 40PercentVsh3PercentPhi-G...
- NSR=1600m-Cum.Gas
- 40PercentVsh3PercentPhi-Phi
- IC4-nC4Ratio

Plot: Summary of ICE Plot Impact for all Features

38% above avg
30% below avg