DATA GATHERING TO WELL PREDICTIONS: UTILIZING A MACHINE LEARNING CASE STUDY FROM THE SPIRT RIVER FORMATION

John Hirschmiller, P.Geo.
PROJECT SCOPE

• Why Machine Learning?
  – Understand and interpret multi-variable problems to optimize well production performance.

• Our Project:
  – Develop a Machine Learning Model to predict 12-month cumulative gas for the Spirit River Formation.
  – ~900 Horizontal wells with production and publicly available completion data.
  – Incorporate multiple disciplines (Geoscience, Reservoir Engineering, Completions) to understand the full story.
OUTLINE

• Spirit River Play/Geological Overview

• Data Organization & Management

• Feature Selection & Model Building

• Data Visualization, Interpretation Tools & Making Predictions
SPIRIT RIVER PLAY/GEOLOGICAL OVERVIEW
GLJ’S SPIRIT RIVER REGIONAL SUBSURFACE ANALYSIS

• GLJ has evaluated significant portions of the Spirit River across all the development areas (grey lands)
SPIRIT RIVER NOMENCLATURE AND ZONES

- 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.

Stratigraphy after Jackson, 1984 & Petrel Robertson 2013
Deposition of the Machine Learning target formations (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.
SPIRIT RIVER PETROPHYSICAL ANALYSIS

• Over 5000 wells picked for formation tops

• Core 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 completed across the trend
KAKWA (NORTH) TO EDSON (SOUTH)

Similar Pay Thickness & Effective Porosity

Falher H
Wilrich A
Wilrich B
Wilrich C

South - EDSON
DATA ORGANIZATION & MANAGEMENT
DATA REQUIREMENTS

• Regionally Consistent Methodology and Data
  – Stratigraphically Consistent.
  – Volumetrically Consistent (ie. Same Petrophysical approach across the play).
  – Completion data in the SAME format.

• DATABASE, DATEBASE and DATABASE
  – Every well NEEDS to have value assigned to it for EACH feature.
    • Needed to train the model.
  – Every location NEEDS to have value assigned to it for EACH feature.
    • Needed to make predictions.

• Projects are usually a 80/20% time split for Data Gathering/Model Building
  – This does NOT include building a regional subsurface database.
  – Building a regional subsurface database can take MONTHS to YEARS.
DATA MANAGEMENT METHODOLOGY

• We have adopted a map-based data storage.
• Pros: Able to update models and make new predictions quickly.
• This allows us to bin data into **two categories**.

**Controllable**
(Inputs vary by well design)

• Completions – **How a well is drilled**
  • Hz Length
  • Frac tonnage
  • Frac spacing
  • Pump rate
  • Fluid type
  • Etc.

**Non – Controllable**
(All inputs are mapped derived)

• Subsurface – **What is in the ground**
  • Zone
  • Geological (Gross, N:G, Phi, Sw etc.)
  • Reservoir Pressure (or gradient)
  • Reservoir Temperature
  • Gas Composition
  • Liquid Yields
  • Etc.

Database of how wells were D&C. Completion design(s) for future wells.

Mapped subsurface parameters for all features in the dataset.
WELL ZONE PLACEMENT – MONTNEY EXAMPLE

- Calculating the target bench individually for each well would be a laborious task.
We have two well pads in multiple benches. But we need an easy way to quantify what zone they are in.

Utilizing deviation surveys and structure maps, we can use GIS tools to help bulk calculate thousands of well zones simultaneously.
By having well zones characterized, we can apply reservoir parameters to the specific zones.

- **Geological**
  - Gross Thickness, Phi, VSh, Sw, etc.
- **Reservoir**
  - Pressure, Temperature, Gas Comp. etc.
UNDERSTANDING LIQUIDS YIELDS
- DUVERNAY CONDENSATE REPORTING ISSUES

- If public data is sparse or inaccurate, we need to be able to supplement it with other data.

- CGR Mapping incorporates:
  - True vertical depth
  - Source rock analysis
  - Geothermal Gradient
  - Magnetic Surveys
  - Other geologic controls
  - Known condensate yields from producing wells

- We can now use an estimate of CGR for each well and location in the Duvernay to feed into a ML model.
DATA PROBLEMS & POINTS OF CAUTION

• Public Data Limitations
  – Completions data, Liquids reporting issues, etc.
  – Data Cleanup and Vetting is pertinent!
  – **Crap In = Crap Out**

• Spatial Features
  – I.e. Latitude and Longitude.
  – Using a spatial location as a parameter may give false prediction capabilities.
  – ML models will be great at predicting nearby wells (Pad Drilling).
  – When stepping out away into virgin reservoir to predict future wells, prediction results will be severely diminished.
  – Latitude and Longitude should **not** be used as a “Proxy” for geology.

• Redundancy
  – Limit features which can be calculated from other features.
  – Net Pay = Gross Pay * Net to Gross
FEATURE SELECTION & BUILDING THE MODEL
DATA GENERATED FOR SPIRIT RIVER MACHINE LEARNING PROJECT

Target:
- 12-month cumulative gas (corrected for downtime)

Well Location Information:
- X, Y, TVD, Region, Well Bore Azimuth

Offset Drainage:
- Gas, Condensate, Oil and Water Production from the same zone within 400, 800, 1600-meter radii of the well at the time it was drilled

Subsurface Information:
- Geological Information: Geologic Zone, Depositional Facies, Mapped Parameters: Gross Pay, Net Pay, Net to Gross, Volume of Shale, Porosity, Water Saturation, Hydrocarbon Pore Volume, at different Volume of Shale & Porosity Cutoffs – 40% & 3%, 50% & 3%, 50% & 2%)
- Reservoir Fluid Information: Pressure, Percentage Over & Under Pressure, Reservoir Temperature, Temperature Gradient, iC4/nC4 Ratio, Shallow-Cut & Deep-Cut Condensate Yields

Completion Information:
- Fracture Technology, Base Fluid, Energizer, Balls Recovered, Estimated Problem Time, Fluid Volume & Fluid Concentration, Percentage Fluid Recovery, Number of Stages, Stage Spacing, Proppant Size, Proppant Concentration (t/m³), Tonnage Placed by Proppant Type, Total Proppant Intensity (t/m), Proppant Intensity by Proppant Type, (Completed) Horizontal Length

Regionally Consistent

~900 wells
>100 features
Predict 12mo cum. gas

GLJ Petroleum Consultants
It’s not just about how many features are used, but which features are selected.

We selected the most predictive features which also considered:

- Stability
- Usability
- Interpretability
- Buy-in
BUILDING THE MACHINE LEARNING MODEL

...Building the predictive machine learning model.....
We have a trained model, now what?

• Model summary
  – ~900 Wells
  – Target = 12 month cumulative gas (mmcf)
  – 13 Features
    - Pay
    - Vsh
    - Proppant intensity

• Now we put the model through its paces to generate predictions and build interpretive tools.
  – Put the model to work!
Feature grouping helps to craft a meaningful story.
## FEATURE IMPORTANCE USING GROUPING

<table>
<thead>
<tr>
<th>Feature</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>HZLength(m)</td>
<td>0.3</td>
</tr>
<tr>
<td>TotalProppantIntensity(t_m)</td>
<td>0.2</td>
</tr>
<tr>
<td>FracSpacing(m)</td>
<td>0.1</td>
</tr>
<tr>
<td>ProppantConcentration(t_m3)</td>
<td>0.1</td>
</tr>
<tr>
<td>ZoneTVD(m)</td>
<td>0.3</td>
</tr>
<tr>
<td>Zone_LABELS</td>
<td>0.2</td>
</tr>
<tr>
<td>PercentOver-UnderPressure</td>
<td>0.1</td>
</tr>
<tr>
<td>40PercentVsh3PercentPhi-GrossPay(m)</td>
<td>0.1</td>
</tr>
<tr>
<td>40PercentVsh3PercentPhi-VSH</td>
<td>0.1</td>
</tr>
<tr>
<td>40PercentVsh3PercentPhi-Phi</td>
<td>0.1</td>
</tr>
<tr>
<td>40PercentVsh3PercentPhi-NG</td>
<td>0.1</td>
</tr>
<tr>
<td>iC4-nC4Ratio</td>
<td>0.0</td>
</tr>
<tr>
<td>NSR=1600m-Cum.Gas</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Completion = Controllable**

**Subsurface = Non-controllable**

*... but selectable... you can choose where you drill your well*
DATA VISUALIZATION, INTERPRETATION TOOLS & MAKING PREDICTIONS
• Visualizing the data is pertinent to understand and what the model is telling us.
• We have built a platform which allows us to **visualize input data**, as well as have **interpretation tools**.
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…
**INDIVIDUAL CONDITIONAL EXPECTATION (ALL WELLS)**

- **Red X**: each well’s actual Proppant value
- **Blue lines**: Individual well predictions
- **Yellow line**: Average production prediction of all wells for each quantile of the Proppant values *(Partial Dependence Plot)*
- **Dashed-red line**: Average production value (1.2bcf 12mo cum.)
- **Maximum Average Impact** caused by changing Proppant values over the range of values in the dataset
Response to Proppant Intensity varies by geologic zone

Uncentered PDP
(impact in the context of actual values)

Centered PDP
(relative impact, amplifies the shape)

ICE Plot, number of unique grid points: 50, target average = 1238 MMcf

GLJ Petroleum Consultants

Not just one recipe for all zones

Incorporate Geoscience Information!
CONTRIBUTION TO PERFORMANCE FOR ONE WELL, BY FEATURE, IN MMCF

Pay effect is negative
Hz length effect is negative
You’re in a good zone
Proppant value is adding to production performance
Aggregate effect is 6% below the average well
Quantitative reconciliation of feature contributions to well performance

Avg Well Production (1.2bcf)
IFLE Value (mmcf)

GLJ Petroleum Consultants
IFLE – COMPARISON OF 2 PROXIMAL WELLS (SAME ZONE)

Well A (Left) = 1.72 Bcf 12-Month Cum
Well B (Right) = 0.97 Bcf 12-Month Cum

Frac Spacing, 1600m Cum Gas, Gross Pay, Horizontal Length, N:G, Over/Under Pressure, Porosity, Proppant Concentration, Proppant Intensity (t/m), VShale, Zone, TVD, iC4/nC4 Ratio

Equal

A>B

GLJ Petroleum Consultants
• We need to be able to easily generate predictions from a model.
• By building a platform, that allows csv/xlsx import and export, anyone who can gather the features (in excel) can generate predictions.
MAKING PREDICTION – REGIONAL SCALE

• How can we look at the play and understand what 12-month cumulative gas can be by section.

• Let the machine do the work for us!

• Vary parameters by section using regional mapping.

• Generic completion design.
NET PAY – MODEL INPUTS

• “Hybrid” Net Pay Map for thickest Zone in the play

• Predicted Production is not a linear relationship to Net Pay
PREDICTED 12 MONTH CUMULATIVE PRODUCTION MAP

- Non-Controllable
  - Zone
  - Gross Pay
  - N:G
  - Porosity
  - VShale
  - Over Under Pressure
  - iC4/nC4 Ratio
  - TVD
  - Offset Production

- Controllable
  - HZ Length
  - Proppant Intensity (t/m)
  - Frac Spacing
  - Proppant Concentration (t/m³)
• Cannot compare directly as completions and HZ length are different.

• Red production bubbles represent 12-month cumulative production.

• Smaller bubbles are in lower prediction areas, whereas large bubbles are in higher predicted areas.
By increasing proppant intensity from 0.8 t/m to 0.95 t/m, we will see an increase of 0.4 Bcf for 12 Month Cum Gas, for this planned well.
OPTIMIZING WELL COMPLETIONS – EXAMPLE 2

- The proppant intensity of this planned well is too high.

- By decreasing the proppant intensity, a similar 12 Month Cum Gas would be achieved.
CONCLUDING REMARKS

• With machine learning we can:
  – Understand current results, and why wells perform the way they do.
  – Quantify what features contributes to variability in production results.
  – Better optimize future performance, and economics.

• We **need** lots of **consistent** data on a regional basis.
  – Geological
  – Reservoir
  – Well Completion
    - Non-Controllable (Subsurface)
    - Controllable (Well Design)
Since acquiring Verdazo AI in September 2019, GLJ is currently updating our databases to make them usable with Machine Learning for several key resource plays in Canada.

NEXT STEPS FOR GLJ’S MACHINE LEARNING

- Cardium
- Montney
- Clearwater
- Ellerslie
- Viking
- Duvernay
- Swan Hills
- Charlie Lake
- Torquay
- Bakken
- Etc.
Questions?

John Hirschmiller, P.Geo.
GLJ Petroleum Consultants
1-403-266-9423
JHirschmiller@gljpc.com