Multivariate Analysis Using Advanced Probabilistic Techniques for Completion Optimization

Bertrand Groulx, Verdazo Analytics; Jim Gouveia, Rose & Associates, LLP; Don Chenery, Verdazo Analytics
Agenda

1) Introduction (Multivariate Analysis)
2) Objective & Purpose of this Study
3) Hybrid Visual Approach
4) Parallel Coordinates Distribution Patterns
5) Input Optimization Distributions
6) Workflow
7) Studies
8) Conclusions
Introduction: Multivariate Analysis

Multivariate Analysis has evolved to include tools, techniques, technologies and workflows that include:

- **Regression analysis** (i.e. “black box” tools) provides predictive outcomes, often with limited supporting evidence, or supporting evidence that is difficult to understand.

- **Visual tools** that effectively communicate correlations/relationships on less complex data, but fail in increasingly complex data.

- **Correlation analysis** proves difficult on complex data as the correlations tend to be weak and vary over the full range of values.

- **Statistical methods** are unique in their ability to provide insights into non-continuous correlations where upper and lower thresholds exist, but are less effective at providing deterministic measures of an input’s effect on an outcome.

... but all of these are not without their dangers (see Multivariate Analysis: Completion Optimization’s Silver Bullet?)
Objective

- We looked outside oil & gas for visual multivariate analysis techniques that could be combined with the strengths of statistical techniques and deliver accessible, transparent insights.

- Our objective was to provide a visual methodology that could:
  - identify patterns of behavior.
  - fuel discovery through targeted investigations.
  - cultivate an understanding of a completion input parameter’s impact on production performance.
Purpose of this Study

To develop a scalable and repeatable visual analysis approach that:

1) Uses statistical techniques that are readily accessible to a broad audience.

2) Offers data-driven statistical insights with visual nuances that can inform completion modelling first principles and advanced regression analysis.

3) Is suitable for small to very large datasets.

4) Is effective even when all inputs are not available for all wells.
Workflow

1) Selection of a performance measure set
2) Analogue well selection
3) Selection of numerical completion design input parameters
4) **Parallel Coordinates Distributions (PCD): input parameter impact analysis**
5) Evaluation of analogue fitness and subset selection
6) Input Optimization Distributions (IOD): input optimization process
Parallel Coordinates

- **Parallel Coordinates** are a common way of visualizing high-dimensional geometry and analyzing multivariate data.

- Earliest documented uses in the late 1800’s.

- Important applications are in collision avoidance algorithms (1987), process control, and more recently in intrusion detection.

Source: [Wikipedia](https://en.wikipedia.org/wiki/Parallel_coordinates)
Parallel Coordinates applied to Completion Analysis

- Applying the same “brushing technique” to more complex data leads to “over-plotting”.

- This does not yield any discernable insights or obvious relationships.
Hybrid Approach: Parallel Coordinates Distribution (PCD)

1. Production Performance Measure Data
2. Input Parameter Data
3. Production Performance Quartile Grouping

Figure 3
Parallel Coordinates Distribution (PCD) Chart
(Shows distributions of Input Values for each Production Performance Quartile)

Distribution of frac spacing for Top Quartile performing wells*

Distribution of frac spacing for Bottom Quartile performing wells

*Note: the production performance measure used here is 12 month cumulative oil per 100m completed length.
Purpose of Parallel Coordinates Distribution (PCD) Chart

- Identify the inputs that warrant optimization investigation (i.e. to focus your efforts where it counts the most)
- Identify patterns that illustrate the concepts of thresholds and correlation windows.
- Use patterns to determine target values used in the binning of Input Optimization Distribution charts (discussed later in this presentation).
- Identify peculiarities in the patterns that may suggest the analogue is not well defined and subsets should be investigated.
PCD Chart Patterns: Clear Performance Pattern

![Diagram showing clear progression from Bottom Quartile to Top Quartile](image)

Data provided by IHS Information Hub and Canadian Discovery VCFD - Dec 16, 2016, 10:31 AM

Figure 4

- >= 0 to < 26 (320)
- >= 26 to < 50 (319)
- >= 50 to < 75 (320)
- >= 75 to < 100 (320)
PCD Chart Patterns: **Threshold**

Below the threshold the Top Quartile shows higher input values. Above the threshold the Bottom Quartile shows higher input values.
PCD Chart Patterns: Correlation Window

Strongest pattern is identifiable between a range of values.

Figure 4
PCD Chart Patterns: No Discernable Pattern

No discernable pattern. At lower values the distributions overlap. At higher values there is no clear progression.

Figure 4
PCD Matrix

- Columns = Production Performance Measures
- Rows = Inputs
- Used to identify which inputs have the greatest impact on production performance and warrant further optimization analysis
- 12 month cumulative production/100m was the performance measure most effective at producing discernable patterns

Performance Measures Used for PCD Quartiles
Identifiable Patterns in Small Datasets
(note similar threshold values in each dataset)

Frac Spacing Distributions for each Production Quartile (80 wells)

Figure 6
Important PCD Considerations

1) Dimensionally normalized performance measures show patterns more clearly

2) Length is the easiest parameter to correct for (e.g. production/100m)

3) Patterns can still be seen in sample sets as small as 80 wells

4) Consider that uncertainty of the mean (and of how representative the distribution is) increases as the sample size get smaller
Workflow

1) Selection of a performance measure set
2) Analogue well selection
3) Selection of numerical completion design input parameters
4) Parallel Coordinates Distributions (PCD): input parameter impact analysis
5) Evaluation of analogue fitness and subset selection
6) Input Optimization Distributions (IOD): input optimization process
Evaluation of analogue fitness and subset selection

**Full Dataset**

*Pattern discrepancy*

Unit production performance increases above 2000m ????

**Subset (extreme proppant density wells excluded)**

Excluding extreme proppant density wells corrected the pattern discrepancy

---

**Figure B-3**

*SPE-185077-MS • Multivariate Analysis Using Advanced Probabilistic Techniques for Completion Optimization • B. Groulx*
Workflow

1) Selection of a performance measure set
2) Analogue well selection
3) Selection of numerical completion design input parameters
4) Parallel Coordinates Distributions (PCD): input parameter impact analysis
5) Evaluation of analogue fitness and subset selection
6) Input Optimization Distributions (IOD): input optimization process
Input Optimization Distribution (IOD) Chart

• Motive is to identify the optimal Input Parameter value.

• Distribution of a Production Performance measure binned by an Input Parameter.

• Use the insights from the PCD patterns to identify:
  a) bottom threshold bin value.
  b) bin size (to highlight a particular threshold value).
  c) top threshold value.
  d) adjusting all of the above to get reasonable well counts in each bin and a visually manageable number of bins.

• Isolate specific bins of interest in an analogue subset, and use smaller bin values to draw more detailed conclusions.
IOD applied to a PCD with a Clear Performance Pattern

**Parallel Coordinate Distribution (PCD) Chart**

Above 70 m the poorer performing quartiles have higher Frac Spacing (i.e. the top quartile has lower Frac Spacing)

**Input Optimization Distribution (IOD) Chart**

As Frac Spacing values (i.e bins) get higher, or lower, than 70 to 80 m, the production performance decreases.

Note: colours on each chart represent different groupings
IOD applied to a PCD with a Threshold Pattern

**Parallel Coordinate Distribution (PCD) Chart**

- Threshold
- Shows a Threshold = 31 (t/100m)
  - Below this value the Top Quartile shows higher values.
  - Above this value the Bottom Quartile shows higher values.

**Input Optimization Distribution (IOD) Chart**

- Production Performance Grouped by Proppant Density (t/100m)
- Top performing distribution has values that correspond to the threshold identified in the PCD chart.

---

Note: colours on each chart represent different groupings
IOD applied to a PCD with a Correlation Window

Parallel Coordinate Distribution (PCD) Chart

Input Optimization Distribution (IOD) Chart

Between a pump rate of 6 and 8.5, the top quartiles show a higher pump rate than the bottom quartiles.

Above 6, production performance increases with increased pump rate up to 8.5. Above 8.5 production performance drops.

Note: colours on each chart represent different groupings
IOD applied to a PCD with No Discernable Pattern

Parallel Coordinate Distribution (PCD) Chart

There is no readily discernable pattern

Input Optimization Distribution (IOD) Chart

There is no readily discernable pattern

Note: colours on each chart represent different groupings
Studies

- Datasets range from 236 wells to 1279 wells
- Studies were performed to illustrate concepts
- A more rigorous analysis should incorporate geological information to refine analogue subsets
Workflow

1) Selection of a performance measure set
2) Analogue well selection
3) Selection of numerical completion design input parameters
4) Parallel Coordinates Distributions (PCD): input parameter impact analysis
5) Evaluation of analogue fitness and subset selection
6) Input Optimization Distributions (IOD): input optimization process
Cardium Oil Study Area

- Tighter Frac Spacing improves unit production performance until Frac Spacing drops below a threshold of ~70 m (see slide 23).
- Increasing Proppant Density improves unit production performance until the upper threshold of ~31 t/100m is reached (see slide 24).
- As Pump Rates increase from 6 m³/min, unit production performance increases up to an upper threshold of ~8.5 m³/min (see slide 25).
When the Production Performance measure is not dimensionally normalized the main insight is that the top quartile wells are longer than the other quartiles.
When the Production Performance measure is dimensionally normalized the main insight is that unit production performance begins to decrease above a ~1400m threshold.
Montney Dry Gas Study Area

- Demonstrated a contradictory PCD pattern on Completed Length that required investigating an analogue subset selection (to separate out extreme proppant density wells).
- Reduction in unit production performance above upper threshold of ~ 2000m completed length.
- Dramatic increase in unit production performance with higher Proppant Density (t/100m) with no upper threshold.

452 wells within the study area:
- Formation = Montney
- Primary Product = Gas
- Cased
- Base Fluid = Slickwater
- Horizontal
- Production Year >2009
- Frac information available
- > 12 months of production

SPE-185077-MS • Multivariate Analysis Using Advanced Probabilistic Techniques for Completion Optimization • B. Groulx
Montney Dry Gas PCD Chart for 12 Month Cumulative Gas (Mcf)
Montney Dry Gas → Analogue Subset Selection

**Full Dataset**

Unit production performance increases above 2000m?

**Subset (extreme proppant density wells excluded)**

Unit production performance is better below 2100m

Note: colours on each chart represent different groupings
Montney Dry Gas IOD Chart of Proppant Density

**Parallel Coordinate Distribution (PCD) Chart**

<table>
<thead>
<tr>
<th>Proppant Density (Tonnes per 100m Completed Length)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>0%</td>
</tr>
</tbody>
</table>

Note: Colours on each chart represent different groupings.

Figure B-4

**Input Optimization Distribution (IOD) Chart**

No upper threshold on Proppant Density’s impact on production performance.

Figure B-4

SPE-185077-MS • Multivariate Analysis Using Advanced Probabilistic Techniques for Completion Optimization • B. Groulx
Montney Liquids-Rich Gas Study Area

- Increased Completed Length shows a reduction in unit production performance (mcf/100m) across the entire range of values
- Proppant Density shows a reduction of unit production performance above ~105 t/100m
- Higher Pumping Rates show an increase in unit production performance with no upper threshold

236 wells within the study area:
- Formation = Montney
- Primary Product = Gas
- Open
- Base Fluid Group = Water
- Horizontal
- Production Year >2008
- Frac information available
- > 12 months of production

Figure C-1
Montney Liquids-Rich Gas PCD Chart for 12 Month Gas (Mcf/100m)

Unit production is better in shorter wells.

Figure C-3
Montney Liquids-Rich Gas IOD Chart of Proppant Density

Parallel Coordinate Distribution (PCD) Chart

Input Optimization Distribution (IOD) Chart

Threshold = 105 (t/100m)

Note: colours on each chart represent different groupings
Montney Liquids-Rich Gas IOD Chart of Pumping Rate

Parallel Coordinate Distribution (PCD) Chart

Best unit production in wells with highest Pumping Rate.

Note: colours on each chart represent different groupings

SPE-185077-MS • Multivariate Analysis Using Advanced Probabilistic Techniques for Completion Optimization • B. Groulx
Conclusions Part 1

- The PCD and IOD visual analysis methodology is suitable for testing any input’s impact on any performance measure.
- PCD Patterns were identifiable in all study areas, suggesting this methodology is suitable for all play types.
- Where rate restriction practices are being used it is recommended that cumulative production (for >= 12 months) be used (i.e. IP90 or other near-term performance measures could yield misleading results).
- Patterns are better defined when performance measures and inputs use dimensional normalization.
- Lateral length is the easiest, and most effective, input parameter to correct for.
Conclusions Part 2

• Inputs require enough statistical variability to see PCD Patterns (observations indicate that P10:P90 of input parameter values need to be >1.6)

• This approach is effective at communicating nuances in the data, such as thresholds and correlation windows.

• Specific threshold values and correlation window ranges can be valuable inputs to other regression techniques or modelling efforts.

• The approach is scalable, accommodating datasets containing greater than 1000 wells down to as few as 80 wells.

• Inexplicable patterns are an effective way of identifying the need for analogue review and subset selection.
The authors wish to acknowledge:

- IHS Markit’s Canadian Information Hub for production data.
- Canadian Discovery Ltd.’s Well Completion and Frac Database (now owned by geoLOGIC Systems) for completion data.
- Verdazo Analytics Inc. and Rose and Associates for their support.
"Why did you not use production/stage as your performance measure?"

Stages work collectively as a system, not independently. As you increase stage spacing you increase the degree to which the stimulated rock volumes for each stage overlap (as evidenced in this image).

Optimal production for a well occurs when the amount of overlap for a collection of stages produces the highest production levels. The maximum production/stage will not deliver optimal well production unless the stages do not interfere with one another.
Supporting slide for anticipated question:

“Why did you choose 12 month cumulative production as your production performance measure instead of EUR?”

The next slide (from the blog [How useful are IP30, IP60, IP90 … initial production measures?](#)) shows the correlations of different production performance measures to EUR.

12 month cumulative production:

- is the minimum number of months required to ensure a strong correlation to EUR
- has enough production to matter, while short enough to include as many recent wells as possible
- is accessible to anyone doing analysis → whereas the only practical way to use EUR would be to rely on auto-forecasting methods which are not readily available to everyone
## 4 Play Analysis of the Correlation of Production Measures to EUR using VISAGE

<table>
<thead>
<tr>
<th></th>
<th>Montney (Gas)</th>
<th>Cardium (Oil)</th>
<th>Viking (Oil)</th>
<th>Bakken (Oil)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data Set 1</td>
<td>Data Set 2</td>
<td>Data Set 1</td>
<td>Data Set 2</td>
</tr>
<tr>
<td></td>
<td>Correlation %</td>
<td>Correlation %</td>
<td>Correlation %</td>
<td>Correlation %</td>
</tr>
<tr>
<td>PD Rate (month 1)</td>
<td>10.6</td>
<td>385</td>
<td>18.9</td>
<td>227</td>
</tr>
<tr>
<td>PD Rate (month 1-2)</td>
<td>21.0</td>
<td>384</td>
<td>29.9</td>
<td>226</td>
</tr>
<tr>
<td>PD Rate (month 1-3)</td>
<td>31.2</td>
<td>383</td>
<td>36.7</td>
<td>223</td>
</tr>
<tr>
<td>Peak</td>
<td>60.0</td>
<td>585</td>
<td>50.6</td>
<td>227</td>
</tr>
<tr>
<td>IP30</td>
<td>32.6</td>
<td>585</td>
<td>39.3</td>
<td>227</td>
</tr>
<tr>
<td>IP60</td>
<td>42.7</td>
<td>585</td>
<td>45.2</td>
<td>227</td>
</tr>
<tr>
<td>IP90</td>
<td>49.2</td>
<td>585</td>
<td>49.9</td>
<td>227</td>
</tr>
<tr>
<td>IP180</td>
<td>60.8</td>
<td>576</td>
<td>62.0</td>
<td>227</td>
</tr>
<tr>
<td>IP365</td>
<td>72.4</td>
<td>576</td>
<td>74.9</td>
<td>227</td>
</tr>
</tbody>
</table>

### Condensed Data

<table>
<thead>
<tr>
<th></th>
<th>Data Set 1</th>
<th>Data Set 2</th>
<th>Data Set 1</th>
<th>Data Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation %</td>
<td>Correlation %</td>
<td>Correlation %</td>
<td>Correlation %</td>
</tr>
<tr>
<td>3 Month Cum</td>
<td>23.2</td>
<td>585</td>
<td>19.4</td>
<td>227</td>
</tr>
<tr>
<td>6 Month Cum</td>
<td>49.3</td>
<td>585</td>
<td>45.1</td>
<td>227</td>
</tr>
<tr>
<td>12 Month Cum</td>
<td>67.1</td>
<td>523</td>
<td>67.0</td>
<td>227</td>
</tr>
<tr>
<td>18 Month Cum</td>
<td>75.4</td>
<td>413</td>
<td>76.1</td>
<td>227</td>
</tr>
<tr>
<td>24 Month Cum</td>
<td>79.7</td>
<td>377</td>
<td>81.6</td>
<td>227</td>
</tr>
<tr>
<td>30 Month Cum</td>
<td>83.5</td>
<td>287</td>
<td>85.1</td>
<td>227</td>
</tr>
<tr>
<td>36 Month Cum</td>
<td>87.5</td>
<td>227</td>
<td>87.5</td>
<td>227</td>
</tr>
<tr>
<td>Non-condensed Data</td>
<td>16.4</td>
<td>585</td>
<td>8.9</td>
<td>227</td>
</tr>
<tr>
<td>6 Month Cum</td>
<td>40.3</td>
<td>585</td>
<td>30.5</td>
<td>227</td>
</tr>
<tr>
<td>12 Month Cum</td>
<td>59.5</td>
<td>523</td>
<td>56.2</td>
<td>227</td>
</tr>
<tr>
<td>18 Month Cum</td>
<td>71.5</td>
<td>413</td>
<td>70.5</td>
<td>227</td>
</tr>
<tr>
<td>24 Month Cum</td>
<td>77.5</td>
<td>377</td>
<td>78.4</td>
<td>227</td>
</tr>
<tr>
<td>30 Month Cum</td>
<td>82.0</td>
<td>287</td>
<td>83.5</td>
<td>227</td>
</tr>
<tr>
<td>36 Month Cum</td>
<td>86.4</td>
<td>227</td>
<td>86.4</td>
<td>227</td>
</tr>
</tbody>
</table>

### Legend

- **Green** = Correlation between 70% and 100%
- **Yellow** = Correlation between 50% and 70%
- **Red** = Correlation between 30% and 50%
- **Grey** = Correlation between 0% and 30%

**Data Set 1** = wells with >80% correlation on Modified Duong fits for both "a" and "m" and >6 months production after peak

**Data Set 2** = subset of Data Set 1 where all wells have >=36 months production

Note: Sample sets include only horizontal wells. [www.visageinfo.com](http://www.visageinfo.com)

EUR calculation based on 240 month forecast using Modified Duong auto-forecast up to boundary dominated flow BDF, then transitioning to Arps for remainder of forecast.

Gas wells (Montney) used 60 months to BDF and a b value of 0.5 for Arps

Oil wells (Cardium, Viking and Bakken) used 48 months to BDF and a b value of 0.5 for Arps