Quantitative Characterization of Fracture Frequency Variations Using a Linear Piecewise Regression Analysis and the Akaike Information Criterion*

Alex P. O’Hara¹ and Robert D. Jacobi¹,²

Search and Discovery Article #42147 (2017)**
Posted November 6, 2017

*Adapted from oral presentation given at AAPG Eastern Section 46th Annual Meeting, Morgantown, West Virginia, September 24-27, 2017
**Datapages © 2017 Serial rights given by author. For all other rights contact author directly.

¹University at Buffalo, Buffalo, New York (aohara999@gmail.com)
²EQT Production, Pittsburg, Pennsylvania

Abstract

We present a new quantitative approach for characterizing fracture frequency variations using a linear piecewise regression (LPR) analysis and the Akaike Information Criterion (AIC). Break points calculated for the LPRs produce linear segments with varying slopes for a cumulative fracture frequency (CFF) curve. An AIC value is calculated for each LPR model in order to determine the optimal number of linear segments that fit the CFF data. The optimal number of segments is obtained by minimizing the AIC value for a single dataset. Results from the statistical analysis produced three CFF slope intervals that define the distribution of possible fracture frequencies unique to the geologic setting from which they were derived. A total of 3678 fracture and vein measurements were collected using scanline, scangrid, and abbreviated methods at 38 sites in the Utica black shale and overlying coarser clastics of the Mohawk Valley in eastern New York State.

To produce a CFF curve, fracture frequency is summed along a transect perpendicular to the strike of the fracture set. The piecewise function in the R package, "Segmented", calculates break points where the slope of the CFF changes. The AIC model selection method produces LPRs with the optimal number of breakpoints and segments by penalizing additional parameters introduced with each new segment. A comparison with the Bayesian Information Criterion (BIC) found that AIC models outperformed the BIC method because the BIC equation over-penalized additional parameters. Segmenting the CFFs produced three unique slope intervals, each with a set of defining characteristics. Background frequencies are defined by an average CFF slope of 8 with no significant changes in slope (including prominent frequency peaks). The average background fracture frequency is 2.4 fractures/m. Transition frequencies exhibit higher CFF slopes, averaging 111, and higher average fracture frequency of 12.3 fractures/m. Fracture intensification domains (including fractures in fault damage zones) are defined by the highest average CFF slope of 1649, produce prominent frequency peaks (>50 fractures/m) and have the highest average fracture frequency of 44.6 fractures/m. Results of the piecewise analysis provide quantified boundaries that can be used to create a fracture frequency framework for a defined geologic setting, aiding in predictions of fracture frequency variations due to local structural features.
Selected References


O’Hara, Alex P., Robert D. Jacovi, and H. David Sheets, 2017, Predicting the width and average fracture frequency of damage zones using a partial least squares statistical analysis: Implications for fault zone development: Journal Structural Geology, v. 98, p. 38-52.
Quantitative characterization of fracture frequency variations using a linear piecewise regression and the Akaike Information Criterion

Alex P. O’Hara\textsuperscript{a} and Robert D. Jacobi\textsuperscript{a,b}

\textsuperscript{a}UB Rock Fracture Group, Department of Geology, University at Buffalo, Buffalo, NY

\textsuperscript{b}EQT Production, EQT Plaza, 625 Liberty Avenue, Pittsburgh, PA
Acknowledgments

- Dr. Charles Mitchell
- Dr. Marcus Bursik
- Dr. David Sheets
- Anna Hrywnak
- Steve Saboda
- Tayler Schweigel
- Richard Nyahay
- John Ebel
- Christopher Willan
- Craig Eckert
- Joe Morris
- Travis Warner
- Scott McCallum
- Sandra Cronauer
Problems in fault damage zone research

- Subjectivity (Choi et al., 2016)
- Defining “background fracture frequency”
- Error introduced in cross-study comparisons

Solutions to Subjectivity

- Quantitative determination of fault damage zone width
- Methodology using piecewise analysis and the Akaike Information Criterion
Application of Quantitative Analysis

Predicting fault damage zone width and average fracture frequency

O’Hara et al., 2017
Importance in developing quantitative analyses of fracture frequency distributions

- Methodology for reliably defining fracture/damage zone boundaries
- Produce consistent damage zone width values for multivariate statistical analyses
- Predictable fracture frequency distributions among varying geologic settings

New hypothesis presented in discussion
Study Area

- **Background**: Modified from O’Hara et al., 2017
- **Methods**
- **Results**
- **Discussion**
- **Conclusions**

The map shows a detailed geological map of the study area with various rock formations and structural features. Notably, the Precambrian formation is highlighted in the center of the map, with the Utica Group and other Ordovician formations shown in different colors and patterns. The Mohawk River is marked with a blue line, and fault trends are indicated with red lines. Outcrops are marked with black circles.

The map provides a comprehensive view of the geological structure and stratigraphy of the study area, essential for understanding the geological history and potential resources.
Study Area: Chronostratigraphy

After Jacobi and Mitchell, 2002
From Mitchell et al., 2006
Study Area

Modified from O’Hara et al., 2017

n=1586
Fracture Intensification Domain (FID)

- High-frequency fracture zone
- Do not pre-suppose fault influence or primary slip surface location
- Can be considered a fracture dominated subset of fault damage zones in specific cases (fault(s) present)
Fracture Intensification Domain (FID)

Map View

Background
Methods
Results
Discussion
Conclusions
Field Methods

[Image of field methods diagram]

- Scanline Orientation
- Fracture Orientation
- True Spacing
- Scanline Spacing

Angle Between Scanline and Fracture: $A$
Scanline Spacing: $SS$
True Spacing: $TS$

$\sin(A)SS = TS$
Field Methods
Cumulative Fracture Frequency (CFF)

fracture spacing (m) fracture frequency = \frac{1}{\text{fracture spacing (m)}}

Background            Methods             Results             Discussion             Conclusions
Piecewise Regression

A: Linear model
   AIC = 319

B: Piecewise model
   AIC = 199
High Bias

[Graph showing a scatter plot with distance (m) on the x-axis and fracture frequency (m⁻¹) on the y-axis. A trend line is also visible, indicating a decrease in fracture frequency as distance increases.]
High Variance

![Graph showing fracture frequency vs. distance](image)

- **Background**
- **Methods**
- **Results**
- **Discussion**
- **Conclusions**
Optimizing Model Complexity

Scott Fortmann-Roe, 2012
Optimum Model Complexity

![Graph showing the relationship between fracture frequency and distance.](Image)
Akaike (AIC) - Bayesian (BIC) Information Criterion

AIC = \( n + n \log(2\pi) + n \log(\text{RSS}/n) + 2(p + 1) \)

\( \text{constant} \quad \text{residual sum of squares} \quad \text{# of parameters} \)

BIC = \( n + n \log(2\pi) + n \log(\text{RSS}/n) + (\log n)(p + 1) \)

- Minimize AIC/BIC values among potential models
- Use change in AIC/BIC values between 2 models
- Significant change in AIC/BIC values ≥ 2
Information Criterion

![Graph showing Δ Information Criterion vs. number of segments]

- **AIC** (Diamonds)
- **BIC** (Squares)

- Significant model improvement
- Non-significant model improvement
- No model improvement
Model Selection

% Background

Methods

% Results

Discussion

% Conclusions
Model Selection

[Graph showing model selection criteria with AIC and BIC values for different numbers of segments.]

**Background**

**Methods**

**Results**

**Discussion**

**Conclusions**
Model Selection

**Methods**

- **Model Selection**

  - **Significant model improvement**
  - **Non-significant model improvement**
  - **No model improvement**

**Results**

- **Graphs**
  - Segment 1: AIC: 515, BIC: 520
  - Segment 2: AIC: 506, BIC: 514
  - Segment 3: AIC: 422, BIC: 433
  - Segment 4: AIC: 414, BIC: 429
  - Segment 5: AIC: 404, BIC: 421
  - Segment 6: AIC: 397, BIC: 408
  - Segment 7: AIC: 384, BIC: 408
  - Segment 8: AIC: 380, BIC: 408
  - Segment 9: AIC: 385, BIC: 416

**Discussion**

- Analysis of model selection criteria.

**Conclusions**

- Summary of findings and implications.
Model Selection

- **Background**
- **Methods**
- **Results**
- **Discussion**
- **Conclusions**

**Model Selection**

![Graph showing model selection criteria](image)

- Significant model improvement
- Non-significant model improvement
- No model improvement

![Graphs showing CFF vs. distance for different segments](image)
Model Selection

**Background**

**Methods**

**Results**

**Discussion**

**Conclusions**
Model Selection

BIC not significant
Model Selection

- **Background**
- **Methods**
- **Results**
- **Discussion**
- **Conclusions**

**Model Selection**

- **Model Selection**
  - **Significant model improvement**
  - **Non-significant model improvement**
  - **No model improvement**

- **Graph**
  - **Δ Information criterion**
  - **Number of segments**
  - **AIC**
  - **BIC**

- **Graphs**
  - **Segments**
  - **Distance (m)**
  - **CFF (m^3)**
  - **AIC**
  - **BIC**

- **Discussion**

- **Conclusions**
Model Selection

AIC not significant
Model Selection

Breakpoints checked against true number of faults observed in outcrop
Fracture Frequency variations

Background
Methods
Results
Discussion
Conclusions
Fracture Frequency variations

Background

Methods

Results

Discussion

Conclusions
Fracture Frequency variations

**Background**

**Methods**

**Results**

**Discussion**

**Conclusions**

![Graph A](image1.png)  
Fracture frequency vs. distance, showing two distinct data sets labeled 347 and 2001. The trend line is straight, indicating a linear relationship.

![Graph B](image2.png)  
Fracture frequency vs. distance, showing a log-log plot with a trend line described by the equation $y = 13.757x^{-0.7063}$ and $R^2 = 0.7131$. The data points are scattered but closely follow the trend line.
Fracture Frequency variations

Background
Methods
Results
Discussion
Conclusions
Cumulative fracture frequency (CFF) segment comparison

[Graph showing cumulative fracture frequency (CFF) over distance (m).
- The graph compares FID and Background segments.
- Key data points include:
  - CFF (m⁻¹)
  - Distance (m)
  - 347
  - 7.85
  - 2001

The graph illustrates the difference in CFF between FID and Background segments, highlighting significant changes at specific distances.]
Cumulative fracture frequency (CFF) segment comparison
Fracture Frequency Intervals

Based on 15 outcrops

- average linear regression
- 1 sigma
- CFF slope segments
### Fracture Frequency Intervals

<table>
<thead>
<tr>
<th>background slope</th>
<th>background y intercept</th>
<th>transition slope</th>
<th>transition y intercept</th>
<th>FID slope</th>
<th>FID y intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9065</td>
<td>1.8498</td>
<td>73.169</td>
<td>3.7616</td>
<td>475.18</td>
<td>71.74</td>
</tr>
<tr>
<td>1.255</td>
<td>3.015</td>
<td>106.23</td>
<td>-1.3595</td>
<td>502.61</td>
<td>-4.3833</td>
</tr>
<tr>
<td>1.759</td>
<td>0.7198</td>
<td>108.52</td>
<td>0.1476</td>
<td>1189.2</td>
<td>-5.0608</td>
</tr>
<tr>
<td>2.1351</td>
<td>-0.5716</td>
<td>111.33</td>
<td>1.2871</td>
<td>1691</td>
<td>-24.602</td>
</tr>
<tr>
<td>2.373</td>
<td>0.2051</td>
<td>112.69</td>
<td>-8.8958</td>
<td>2090.4</td>
<td>61.267</td>
</tr>
<tr>
<td>3.3435</td>
<td>4.4701</td>
<td>146.12</td>
<td>-13.934</td>
<td>2301.1</td>
<td>135.1</td>
</tr>
<tr>
<td>4.4894</td>
<td>-2.876</td>
<td></td>
<td></td>
<td>3014.400</td>
<td>155.25</td>
</tr>
<tr>
<td>5.1641</td>
<td>0.0024</td>
<td></td>
<td></td>
<td>3409.8</td>
<td>24.08</td>
</tr>
<tr>
<td>5.384</td>
<td>-0.1425</td>
<td></td>
<td></td>
<td>5200</td>
<td>33.542</td>
</tr>
<tr>
<td>6.7791</td>
<td>-1.2958</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.772</td>
<td>-1.0865</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.403</td>
<td>-0.2261</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>avg</strong></td>
<td><strong>4.564</strong></td>
<td><strong>109.677</strong></td>
<td><strong>-3.166</strong></td>
<td><strong>2208.188</strong></td>
<td><strong>49.659</strong></td>
</tr>
<tr>
<td><strong>std dev</strong></td>
<td><strong>3.431</strong></td>
<td><strong>23.177</strong></td>
<td><strong>6.795</strong></td>
<td><strong>1511.846</strong></td>
<td><strong>62.665</strong></td>
</tr>
<tr>
<td><strong>avg+std dev</strong></td>
<td><strong>7.995</strong></td>
<td><strong>132.854</strong></td>
<td><strong>3.629</strong></td>
<td><strong>3720.034</strong></td>
<td><strong>112.324</strong></td>
</tr>
<tr>
<td><strong>avg-stdev</strong></td>
<td><strong>1.133</strong></td>
<td><strong>86.499</strong></td>
<td><strong>-9.960</strong></td>
<td><strong>696.342</strong></td>
<td><strong>-13.006</strong></td>
</tr>
<tr>
<td><strong>std error</strong></td>
<td><strong>0.990</strong></td>
<td><strong>9.462</strong></td>
<td><strong>2.774</strong></td>
<td><strong>503.949</strong></td>
<td><strong>20.888</strong></td>
</tr>
<tr>
<td><strong>ci+</strong></td>
<td><strong>6.505</strong></td>
<td><strong>128.222</strong></td>
<td><strong>1.460</strong></td>
<td><strong>3195.927</strong></td>
<td><strong>90.601</strong></td>
</tr>
<tr>
<td><strong>ci-</strong></td>
<td><strong>2.622</strong></td>
<td><strong>91.131</strong></td>
<td><strong>-0.783</strong></td>
<td><strong>1220.448</strong></td>
<td><strong>8.718</strong></td>
</tr>
</tbody>
</table>

**Background**

- Description of the background of the study.

**Methods**

- Details of the methods used in the study.

**Results**

- Presenting the results obtained from the study.

**Discussion**

- Analysis and discussion of the results.

**Conclusions**

- Summary and conclusions drawn from the study.
Fracture Frequency Intervals

[Graph showing CFF (m⁻¹) vs. distance (m) with background, transition, and FID regions.]
Fracture Frequency Intervals

Graph showing the relationship between distance (m) and CFF (m⁻¹) with different lines representing different data sets and their confidence intervals.
Strain Localization

**Background**

Methods

Results

Discussion

Conclusions

**Grain Distribution**

**Fault Damage Zone Boundaries**

---

**Grain Size Distribution ($D_2$)**

$D_2 = \text{grain size/grain density}$

$D_2 \sim 2$ (fault core)

Frost et al., 2009
Strain Localization

A

B

FID core
FID transition
background

high-strain zone
low-strain zone

c

distance (m)

CFF (m⁻¹)

0
0
500
1000
1500

FID development

t₁

t₂

t₃

Map View

strain localization
Fracture variations on a fault

cross section

background fracture set
transition FID
fault tip FID
compound FID
normal slip induced FID

fault / fractures

Background  Methods  Results  Discussion  Conclusions
Conclusions

• Research produced a methodology for reliably defining fracture/damage zone boundaries using linear piecewise regressions and AIC
• Background, transition and FID fracture sets produce unique CFF slope responses
• Predictable fracture frequency distributions among varying geologic settings
• Strain localization controls fracture formation during fault initiation