Version of Record: https://www.sciencedirect.com/science/article/pii/S0191814122000232
Manuscript_6d36787ceb2008f9f419aa721109a479

Multiscale Spatial Analysis of Fracture Arrangement and Pattern Reconstruction using Ripley's K-Function

Mahmood Shakiba [a,*], Larry W. Lake [a], Julia F.W. Gale [b], Michael J. Pyrcz [a,b,c]

[a] Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, TX 78712, USA

[b] Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78713, USA

[c] Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78712, USA

[*] **Corresponding author:**

Mahmood Shakiba

E-mail address: m_shakiba@utexas.edu

Tel: +1 (512)-656-1255

The University of Texas at Austin
Hildebrand Department of Petroleum and Geosystems Engineering
200 E. Dean Keeton St., Stop C0300
Austin, TX 78712-1585, USA

Keywords:

Ripley's K-function

Fracture arrangement

Pattern reconstruction

Spatial clustering/anti-clustering

Point pattern analysis

Multiscale Spatial Analysis of Fracture Arrangement and Pattern Reconstruction using Ripley's *K*-Function

Mahmood Shakiba[1], Larry W. Lake[1], Julia F.W. Gale[2], Michael J. Pyrcz[1,2,3]

1: Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin

2: Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin

3: Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin

Abstract

This work presents novel multiscale spatial data analytics using Ripley's *K*-function, as a measure of spatial interaction, to study one-dimensional arrangement of fractures. Fracture spatial arrangements are classified into clustered, anticlustered, or indistinguishable from random by testing statistical significance of the calculated Ripley's *K*-function. Characterizations of fracture arrangements are performed as a function of length scale and position. Analysis of the *K*-function along the study interval identifies where fracture clustering and anticlustering occur. A simulation technique is also introduced here to statistically reconstruct spatial arrangements and to generate fracture realizations that are spatially similar to the fractures observed in the field. With this simulation technique, one can also fill spatial gaps in fracture measurements where data are absent, unreliable, or unused. Synthetic as well as field-measured 1D fracture datasets are used for testing and demonstration. Methods introduced in this work can be readily applied to fracture datasets observed in outcrops, borehole image logs, and cores.

1. Introduction

When rock fractures are evaluated as a set, their spatial arrangement becomes especially important (e.g., Smith and Schwartz, 1984; Berkowitz, 2002; Laubach, 2003; Olson, 2004; Narr et al., 2006; Gale et al., 2014; Hardebol et al., 2015). Spatial arrangement describes how arrays of objects are

positioned in a space based on the sequence of and the distances between the objects (e.g., Laubach et al., 2018). Subsurface observations show that fracture networks often span multiple scales and may exhibit diverse spatial arrangements (e.g., Gillespie et al., 2001; Gauthier et al., 2002; Hooker et al., 2009; Questiaux et al., 2010; Gale et al., 2014; Laubach et al., 2018; Li et al., 2018; Wang et al., 2019; Bagni et al., 2020; Dichiarante et al., 2020). Fractures may be clustered (e.g., Laubach, 1992; Marrett et al., 2018; Li et al., 2018; Wang et al., 2019), anticlustered — also referred to as *more regularly spaced* — (e.g., Odling et al., 1999; Gillespie et al., 1999), or indistinguishable from random (e.g., Narr and Suppe, 1991). Hence, spatial arrangements may transition from one state to another as the scale of measurement changes (e.g., Gillespie et al., 1999; Sanderson and Peacock, 2019). For example, a meter-scale fracture cluster may appear regularly spaced if inspected by a magnifying lens. These multiscale variations make characterization of fracture arrangements scale dependent.

Spatial arrangements are quantified using the sequence of and the distances between fractures. In 1D applications, the distances between fractures can be calculated by spacing data. Spacing is equivalent to the distance between intersection points of neighboring fractures with a sampling line interval, such as a scanline, borehole, or core, placed perpendicular to the orientation of fractures. Traditional statistical measures that summarize histograms and the size distribution of fracture spacings, such as the mean μ, the standard deviation σ, and the coefficient of variation $C_v = \sigma/\mu$, do not account for spatial arrangements, because they are calculated regardless of the location and the sequence of fractures. To illustrate this issue, consider two fracture sets with identical fracture spacing histograms and coefficients of variation, but with two different spatial arrangements (**Fig. 1**). In one set, fractures are clustered around the middle, whereas in the other set, fractures are randomly distributed. In this example, the mean μ, the standard deviation σ, and

47 the coefficient of variation C_v are unaffected by the change of spacing sequence and yield identical
48 values in both fracture sets.

49

50 **Figure 1: Comparison of spatial arrangements in two fracture sets. (a) Both fracture sets have identical mean**
51 **μ, standard deviation σ, coefficient of variation C_V, and histograms of fracture spacing. (b) However,**
52 **they exhibit different spatial arrangements. Set 2 is clustered, whereas set 1 is indistinguishable from**
53 **a random arrangement.**

54 To identify fracture arrangements, some methods use cumulative distribution functions for fracture
55 spacing (e.g., Sanderson and Peacock, 2019; Bistacchi et al., 2020); however, they perform spatial
56 analysis only at one scale (i.e., the largest scale of measurement). To perform a multiscale spatial
57 analysis, methods such as lacunarity (Roy et al., 2014) and correlation count (Marrett et al., 2018)
58 use incremental sizes of length scale, similar to lag distances in semivariograms (e.g., Laubach et
59 al., 1995; Hanke et al., 2018). However, these methods quantify the average spatial arrangement
60 and provide no information on the variation of fracture arrangement as a function of position.

61 Fracture arrangement may not be the same everywhere along the study interval. For example, if a
62 scanline measurement is divided into different segments, and each segment is analyzed separately,
63 different fracture arrangements may be observed at a given scale (Bistacchi et al., 2020). Fracture
64 arrangement can be clustered in one segment and anticlustered in another segment. Therefore,
65 when fracture arrangement is reported as an average, these local variations are overlooked.
66 Evaluating the changing nature of fracture arrangement as a function of position can objectively
67 reveal additional information about location, boundaries, and properties of clustered and
68 anticlustered structures along a study interval.

69 Another valuable aspect of quantifying spatial arrangement of fractures is that one can statistically
70 reconstruct (simulate) fracture arrangements and generate fracture realizations that are spatially
71 similar to the fractures observed in the field. For ground water and petroleum engineering

subsurface applications, multiple synthetic fracture realizations are commonly generated, such as for uncertainty quantification and for testing production scenarios (e.g., Shakiba, 2014, 2018; Nejadi et al., 2017; Sun et al., 2017). Such fracture realizations should be generated based on the fracture datasets observed in borehole image logs, cores, or outcrop exposures, to ensure that they represent realistic spatial arrangements. Several methods have been introduced in the literature to simulate fracture arrangements. However, they apply only to specific types of spatial arrangements, such as uniform, random, periodic, or fractal arrangements, and they assume the same type of arrangement across all length scales (e.g., Gillespie et al., 1993; Borgos, 2000; Sanderson and Peacock, 2019). As mentioned, fractures may exhibit diverse spatial arrangements that change from one scale to another. Therefore, more advanced and universal methods are required for simulating naturally occurring fracture arrangements.

This work develops novel spatial quantitative tools using Ripley's *K*-function, as a measure of spatial interaction, to characterize, analyze, and model fracture arrangements across various scales. Ripley's *K*-function has been extensively used in various fields, such as ecology, geography, and astronomy to describe spatial arrangement of discrete points or events (e.g., Stoyan and Penttinen, 2000; Stein et al., 2000; Marcon and Puech, 2003; Wiegand and Moloney, 2013). Fractures observed in outcrops, borehole image logs, or subsurface cores are examples of discrete objects that can be analyzed by Ripley's *K*-function.

In this paper, we first describe how Ripley's *K*-function is calculated for 1D fracture measurements and how it can be interpreted to identify statistically significant clustered and anticlustered arrangements. We use a natural fracture dataset for demonstration. Next, we develop a novel approach based on a local version of Ripley's *K*-function to visualize variation of spatial arrangement along the study interval. This approach identifies the location and properties of

fracture clusters and anticlusters along the study interval. Finally, we introduce a simulation technique using Ripley's *K*-function to generate fracture realizations that honor the spatial arrangement of the measured dataset. We indicate how the simulation technique can be deployed to fill in gaps in measurements. Although examples presented in this paper are based on 1D datasets, published 2D and 3D applications of Ripley's *K*-function in other fields allow our modeling results to be readily extended to practical problems in 2D and 3D subsurface fracture prediction.

2. Fracture dataset

The 1D fracture dataset used here is from the work of Marrett et al. (1999, 2018), who performed fracture measurements on a limestone outcrop exposed on the bed of the Pedernales River in Pedernales Falls State Park, Texas (**Fig. 2a**). Here the limestone strata are from the Upper Member of the Pennsylvanian Marble Falls Formation, and they were exposed by erosion through the Glen Rose Limestone found on the slopes on both sides of the river. The limestone rocks are crinoid-rich, gray packstones and wackestones that suggest a shallow-marine setting. Facies comprise algal mounds, chert, siliceous limestone (silica derived from sponge spicules), and shale (Erlich and Coleman, 2005; Gomez, 2007; Wright et al., 2007). Two sets of fractures occur at Pedernales, NNE-striking, mostly unmineralized joints and EW-striking calcite-filled veins. Both sets are sub-perpendicular to bedding planes. Veins at Pedernales are more abundant and cross multiple layers. Some of the veins are cut by stylolites. Spatial analysis here is performed only for the veins (hereinafter fractures) (**Fig. 2b**).

Figure 2: Marble Falls Limestone and veins at Pedernales. (a) Limestone outcrop exposed on the bed of the Pedernales River in Pedernales Falls State Park, Texas. (b) Image of the calcite-filled veins.

Spacings between 916 fractures were obtained along a scanline approximately 59 m long, perpendicular to the orientation of the fractures. Color contrast between fracture-filling calcite cement and the surrounding rock matrix facilitates measurement of fracture spacings. Fracture spacings larger than 5 mm are measured with a ruler or tape with ± 1 mm precision. Fracture spacings less than 5 mm are measured with a logarithmically graduated comparator (Ortega et al., 2006) and with the aid of a 10x magnifying lens. The comparator allows measurement of spacings as small as 0.05 mm. Fractures with aperture sizes or spacings less than 0.05 mm are ignored. The position of a fracture is the distance from the start of the scanline to the fracture wall. Subsequently, spacings are calculated as the distances between positions of two neighboring fractures.

In addition to fracture positions and spacings, aperture sizes were also measured using comparators. However, we used only the position and spacing data to analyze fracture arrangement in this work. For more details on the dataset, measurement techniques, and regional geology, see papers by Marrett et al. (1999, 2018).

Fracture locations (stick plot) and fracture frequency over equally sized bins along the scanline at Pedernales are illustrated in **Fig. 3**. Most of the fractures cluster around the left side of the study interval. There is an abundance of fractures between 0 and 27 m, and a paucity between 27 and 59 m. Between 0 and 27 m, intensity is non-uniform, and fractures form clusters of various sizes (marked by spikes in fracture intensity, for example between 3.8 m and 5.2 m). Such spatial arrangements of inter-nested clusters are qualitatively fractal (Gillespie et al., 1999; Marrett et al., 2018).

Figure 3: Visualization of fracture measurement at Pedernales. Stick plot indicating fracture locations (top), and variation of frequency over 50 equal-size bins of 1.16 m along the study interval (bottom).

Furthermore, the size distribution of fracture spacings spans four orders of magnitude and is approximately lognormal, identified by a Gaussian-shaped histogram on a logarithmic axis (**Fig. 4**). Summary statistics for the size distribution of spacings are in **Table 1**. Based on the estimated values of the sample mean ($\mu = 64$ mm) and the standard deviation ($\sigma = 156$ mm), the coefficient of variation for fracture spacings is calculated as $C_V = \sigma/\mu = 2.43$. A C_V of greater than 1 implies a wide range of spacings, which clustered patterns possess (Gillespie et al., 1999). However, C_v provides no information about how the *arrangement* of fractures changes as a function of scale and along the study interval. Coefficient of variation applies to nearest-neighbor fracture spacings, and it only confirms a strongly clustered arrangement at small scales. Longer range fractal spatial arrangements cannot be discerned from nearest-neighbor spacings without data on the spacing sequence, and the same holds for other aspects of arrangement.

Figure 4: Histogram (bottom) and box plot (top) of fracture spacing at Pedernales. Spacings are on a logarithmic axis. Box plot displays the first and third quartiles (box edges), the median (middle line inside the box), variability outside the first and third quartiles (whiskers), and outliers (red squares) of spacing data.

Table 1:Summary statistics for size distribution of fracture spacings at Pedernales. Units are millimeters.

In the next section, a method of point pattern analysis is introduced to characterize the spatial arrangement of fractures measured at Pedernales over multiple length scales.

3. Ripley's *K*-function

Ripley's *K*-function (hereinafter *K*-function) is a method of point pattern analysis that describes spatial arrangements (Ripley, 1976, 1977; Baddeley et al., 2007; Dixon, 2014; Cressie, 2015). The *K*-function applies to an array of discrete objects, and it quantifies spatial arrangements using the distances between the objects. Fractures sampled along line intervals, such as in outcrop scanline

166 measurements, borehole image logs, or subsurface cores, are examples of discrete objects that can
167 be analyzed by the *K*-function. In a fracture set with semi-parallel fractures, the orthogonal
168 distances between fractures can be reliably calculated by spacing data. In 1D measurements,
169 fracture spacings are measured as the distances between intersection points of neighboring
170 fractures with a sampling line interval. Consequently, the distance between two fractures of the
171 same set equals the cumulative sum of fracture spacings that fall between these two fractures.
172 Therefore, 1D fracture arrangements can be quantified by the *K*-function.

173 The main advantage of the *K*-function is that it takes scale as a primary variable in the spatial
174 analysis. This makes the *K*-function well suited for studying multiscale characteristics of fracture
175 arrangements. The *K*-function is defined as the "expected" number of neighboring points that fall
176 within the distance h from a point randomly selected inside the study domain, normalized by the
177 point intensity.

$$K(h) = \frac{1}{\lambda} \times E[\#\ of\ points\ within\ distance\ h\ from\ a\ randomly\ selected\ point] \qquad (1)$$

178 where $K(h)$ is the *K*-function at the length scale h; $E[\,]$ is the expected value or expectation; and
179 λ is the point intensity — the total number of points divided by the length of the study interval.
180 $K(h)$ has a unit of length and is a measure of spatial dependence or interaction between points.

181 **3.1 One-Dimensional Formulation of Ripley's *K*-function**

182 Formulations of the *K*-function can be derived for 1D, 2D, and 3D applications. Since our analysis
183 deals with 1D fracture datasets, only the 1D formulation is introduced. The 1D formulation of the
184 *K*-function applies to fracture datasets obtained from scanline measurements, borehole image logs,

or subsurface cores. Calculation of the *K*-function is based on a point counting procedure. At every length scale h, a search interval of half-width h is centered on the first fracture, and the total number of neighboring fractures that fall within the search interval is counted. This step is repeated by centering the search interval on all the other fractures and counting the number of neighbors. **Fig. 5** shows a schematic example of the fracture counting procedure. Finally, all the counted numbers of neighbors are averaged and divided by fracture intensity. Therefore, the *K*-function in a 1D application can be estimated by

$$K(h) = \frac{L}{N^2} \sum_{i=1}^{N} \sum_{j=1 \mid j \neq i}^{N} \omega_i \times H\left[h - \left|x_j - x_i\right|\right] \quad (2)$$

where L is the total length of the study interval; N is the total number of fractures; $\left|x_j - x_i\right|$ is the distance between fractures i and j along the study interval; H is the Heaviside step function (which returns 0 for $h - \left|x_j - x_i\right| < 0$, and 1 otherwise); and ω_i is a weighting factor that corrects for edge effects.

Figure 5: Example of fracture counting used in *K*-function calculation.

Note that the 1D formulation of the *K*-function is analogous to the correlation sum (e.g., Du Bernard et al., 2002; Grassberger and Procaccia, 2004; Marrett et al., 2018; Sanderson and Peacock, 2019), except that the *K*-function incorporates an alternative method for edge correction, and it is normalized by the length of the study interval (L). Normalizing by length allows comparison of the *K*-functions calculated for datasets of different lengths or fracture intensities. Since *K*-function analysis is well-established in many other fields, many tools and methods can be

readily deployed for characterization of fracture arrangements, such as methods for edge correction in 2D and 3D, analytical forms for various spatial patterns, and research codes.

3.2 Edge Correction

Edge or boundary effect refers to a condition where part of a search interval lies outside the study domain **(Fig. 6).** Since fractures outside the domain boundaries are not measured, the actual number of neighboring fractures in this situation is unknown. Inability to count neighboring fractures that are outside the study domain results in an underestimation of the *K*-function (i.e., a bias low). This issue becomes more significant at large length scales because it becomes more likely that part of the search interval would extend beyond the study interval.

Figure 6: Illustration of edge effect. Part of the search interval shown on the left falls outside the study domain. Eq. 3 can be used to calculate weighting factor for edge correction. The search interval on the right, however, is completely inside the study domain; therefore, weighting factor is 1.

To correct for edge effects, a weighting factor is used in the *K*-function formulation to increase the number of counted neighboring fractures proportional to the fraction of the search interval that falls outside the study domain. Ripley (1976, 1977) proposed the following equation for the weighting factors.

$$\omega_i = \frac{2h}{d_{inside}} \tag{3}$$

where h is the half-width of the search interval (same as the length scale h in Eq. 2), and d_{inside} is the part of the search interval that overlaps with the study interval. If all the search interval centered on a specific fracture is inside the study domain, the associated weighting factor equals 1 ($\omega_i = 1$), meaning that edge correction is not required.

3.3 Normalized *K*-function

In a random arrangement, the *expected* number of neighboring fractures within the distance h from a randomly selected fracture is calculated by multiplying the size of the search interval ($2h$) by the fracture intensity ($\lambda = N/L$). Therefore, the expected value of $K(h)$ in a random arrangement is

$$K_{random}(h) = \frac{1}{\lambda} E[\#\ of\ fractures\ within\ distance\ h] = \frac{L}{N}\left(\frac{N}{L} 2h\right) = 2h \tag{4}$$

Values of $K(h)$ greater than $K_{random}(h)$ indicate fracture clustering, whereas values of $K(h)$ smaller than $K_{random}(h)$ indicate anticlustering at the length scale h. To simplify comparisons, the calculated values of $K(h)$ are further normalized by $K_{random}(h)$. Thus, at the length scale h, the normalized *K*-function is

$$K_n(h) = \frac{K(h)}{K_{random}(h)} = \frac{L}{2hN^2} \sum_{i=1}^{N} \sum_{j=1\,|\,j\neq i}^{N} \omega_i \times H\left[h - \left|x_j - x_i\right|\right] \tag{5}$$

Therefore, for random arrangements $K_n(h) = 1$. Consequently, $K_n(h) > 1$ shows fracture clustering, whereas $K_n(h) < 1$ shows anticlustering. However, note that not all clustered or anticlustered arrangements are statistically significant. Even in a random arrangement, the *calculated* $K_n(h)$ can still be greater or less than the *expected* value of 1, indicating some degree of clustering or anticlustering. For random arrangements, $K_n(h) = 1$ refers to the average of many realizations of random arrangements (i.e., the expected value). Therefore, a test of statistical significance is needed to identify whether an observed fracture clustering or anticlustering is statistically significant.

4. Statistical Bootstrap

For any statistical measure, quantifying confidence limits and uncertainty is important for understanding the variability and significance of measurements, experiments, or calculations to support interpretation and decision making. For classification of fracture arrangements, variability in random spatial arrangements can be specified by a confidence interval to identify clustered or anticlustered arrangements.

Here the statistical bootstrap is used to test the randomness hypothesis and to calculate uncertainty interval of the *K*-function. The bootstrap method is a statistical technique that involves random sampling with replacement (Efron, 1992; Hesterberg, 2011). The statistical bootstrap applies Monte Carlo simulation, sampling with replacement, to assess uncertainty in any statistics. It assumes stationarity as well as independent and representative sampling; more advanced workflows of spatial bootstrap with declustering are discussed in Pyrcz and Deutsch (2014).

4.1 Confidence Interval for Random Fracture Arrangements

To find the confidence interval for the *K*-function of random arrangements, the statistical bootstrap generates multiple realizations of random fracture arrangements with the same intensity as that of the measured dataset. The *K*-function is then calculated for each realization. The confidence interval serves as a benchmark envelope to identify arrangements significantly different from a random arrangement. Lower and upper limits of the confidence interval are determined by the choice of the confidence level (i.e., the probability for a statistic to be within the bounds, given that the random hypothesis is correct).

Fig. 7 depicts the *K*-function and the normalized *K*-function of the Pedernales fractures. The shaded areas are the 95% confidence intervals for the random fracture arrangements calculated by 100 bootstrap realizations. Values of $K_n(h)$ are greater than 1 at all length scales, indicating a

persistent multiscale clustered arrangement. However, $K_n(h)$ is inside the confidence interval between length scales of 1×10^{-4} and 3×10^{-4} m and, therefore the spatial arrangement is indistinguishable from random, whereas at larger length scales $K_n(h)$ is above the confidence interval, indicating a statistically significant clustered arrangement.

Figure 7: Spatial analysis of fracture arrangement at Pedernales. Stick plot showing fracture locations along scanline (top); calculated *K*-function (left); and normalized *K*-function (right). At length scales of less than 3×10^{-4} m, the fracture arrangement is indistinguishable from random, whereas at larger length scales a statistically significant clustered arrangement appears.

Moreover, the *K*-function and the normalized *K*-function are approximately linear on the log-log plots between 2×10^{-3} and 30 m length scales. This power law behavior implies a fractal geometry (Gillespie et al., 1999; Marrett et al., 2018) in which larger clusters consist of groups of smaller clusters. The degree of fractal clustering can be estimated by the correlation dimension (i.e., the slope on the log-log plot). The correlation dimension from the *K*-function plot is 0.83, which is consistent with the correlation dimension of 0.8 estimated from the correlation sum plot (Marrett et al., 2018).

For comparison, **Fig. 8** shows the *K*-function and the normalized *K*-function calculated for a synthetic random arrangement. The *K*-function and the normalized *K*-function fall within the confidence interval. Although the values of $K_n(h)$ are greater or less than 1 at some length scales, they are not statistically significant.

Figure 8: Spatial analysis of fracture arrangement in a dataset generated by random placement of fractures. Stick plot showing fracture locations along the study interval (top); calculated *K*-function (left); and normalized *K*-function (right). The *K*-function and normalized *K*-function fall inside confidence interval across all length scales.

4.2 Uncertainty in Ripley's *K*-function

As mentioned in section 2, measurement of small spacings in the field is challenging, and it may require additional tools such as a graduated comparator and a hand lens. Because of such limitations, fractures that are spaced less than 0.05 mm were not recorded at Pedernales. Therefore, accuracy of the measurement is expected to degrade at small scales. Departure of the *K*-function from a fractal behavior (the linear segment on the log-log plot) at length scales of less than 1 mm could be caused by the lower measurement accuracy and truncation. To calculate the uncertainty interval in the *K*-function, the statistical bootstrap is deployed.

Fig. 9 depicts the *K*-function and the normalized *K*-function calculated for the Pedernales dataset with the uncertainty intervals. At small length scales, the bootstrap sampling process produces a larger uncertainty interval. Spatial arrangement at these small scales is mostly controlled by only a few fractures tightly clustered around each other. Therefore, sampling with replacement produces a large uncertainty interval. However, as length scale and the number of neighboring fractures increase, the effect of sampling decreases and the uncertainty interval becomes smaller.

Figure 9: Uncertainty intervals in the *K*-function and normalized *K*-function calculated for Pedernales dataset using statistical bootstrap.

5. Local fracture arrangement

The *K*-function discussed in sections 3 and 4 quantifies the overall fracture arrangement at each scale by pooling all fractures across the study interval. This type of analysis quantifies fracture arrangement as a function of scale but provides no information on the changing nature of fracture arrangement along the study interval. Fracture arrangement may change from clustered in one subregion to anticlustered in another subregion.

In this section, we introduce a new approach that uses a local version of the *K*-function to evaluate fracture arrangement as a function of length scale and position simultaneously. This approach visualizes variation of the *K*-function as a function of position at various scales. By studying local variations of fracture arrangement, one can objectively identify the location and size of fracture clusters and anticlusters.

5.1 Local *K*-function

Recall that the *K*-function in Eq. 2 estimates the "average" number of neighboring fractures at each length scale. The inner sum in Eq. 2 counts the neighboring fractures, and the outer sum integrates the same calculation over all the fracture locations. In other words, the final value of the *K*-function in Eq. 2 is the average of all the local values of the *K*-function calculated for each individual fracture. Therefore, the local estimator for the *K*-function at the location of fracture i (referred to as the local *K*-function) at the length scale h can be represented by

$$K^i(h) = \frac{L}{N} \sum_{i=1 \mid i \neq j}^{N} \omega_i \times H\left[h - \left|x_j - x_i\right|\right] \tag{6}$$

Similarly, the normalized local *K*-function will be

$$K_n^i(h) = \frac{L}{2hN} \sum_{i=1 \mid i \neq j}^{N} \omega_i \times H\left[h - \left|x_j - x_i\right|\right] \tag{7}$$

Normalized local *K*-functions calculated for the Pedernales fractures are shown in **Fig. 10**. The *x*-axis is the location along the scanline and the *y*-axis is length scale. Every point in this figure is colored with the associated value of the normalized local *K*-function. Values of $K_n^i(h)$ greater than 1 (warm colors) indicate fracture clustering, whereas values of $K_n^i(h)$ less than 1 (cool colors) indicate anticlustering. Calculations are performed only at the fracture locations. Regions without

fractures or with zero *K*-function are shown in white. Clearly, the left side of the study interval, near the boundary, shows a higher degree of clustering (i.e., relatively greater values of $K_n^i(h)$) than the remainder of the study interval. Also, multiple clusters exist in this region. However, to demonstrate that these clustered and anticlustered arrangements are indeed statistically significant, we applied the statistical bootstrap to individual fractures.

Figure 10: Profile of normalized local *K*-functions calculated for Pedernales fractures. For each fracture along scanline (*x*-axis), normalized local *K*-function is calculated at each length scale (*y*-axis). Points in the profile are colored with the associated values of normalized local *K*-function. Warm colors (values greater than 1) indicate fracture clustering, whereas cool colors (values less than 1) indicate fracture anticlustering.

To calculate the confidence interval for $K_n^i(h)$ at the location of fracture i, all the other $N-1$ fractures are randomly redistributed while holding the location of fracture i fixed. For every new realization j, the normalized local *K*-function at the location of fracture i, $K_n^{i,j}(h)$, is calculated using Eq. 7. We repeated this process several times ($j = 1, \dots, M$) to obtain a statistically large number of samples. Finally, by comparing the observed value of $K_n^i(h)$ with that of the M realizations ($K_n^{i,j}(h)$ for $j = 1, \dots, M$), we find the *p*-value (or percentile) for fracture i at the length scale h (Eq. 8). We repeat this process for all the other fractures ($i = 1, \dots, N$) and across different length scales.

$$p_i(h) = \frac{Rank(K_n^i(h))}{M+1} \tag{8}$$

Fig. 11a is a profile of the *p*-values calculated for the Pedernales fractures based on 99 bootstrap realizations for each fracture ($M = 99$). Fractures with greater *p*-values are located within clusters, whereas fractures with lower *p*-values are located within anticlustered arrangements.

352 **5.2 Fracture Arrangement along the Study Interval**

353 A main advantage of the local analysis of fracture arrangement is that cluster boundaries at each
354 scale can be explicitly identified. To find the clustered and anticlustered arrangements that are
355 statistically significant, the lower and upper bounds for *p*-values should be selected. In this work,
356 the bounds of the *p*-values for statistically significant anticlustered and clustered arrangements are
357 0.025 and 0.975, respectively, for a 95% confidence interval.

358 Final profile of the statistically significant clustered and anticlustered arrangements at Pedernales
359 is plotted in **Fig. 11b**. Red and purple highlight statistically significant clustered and anticlustered
360 arrangements, respectively. Analysis of this figure reveals several important features. First, as
361 expected, the left side of the study interval up to 27 m exhibits statistically significant clustering
362 of fractures, whereas the remainder of the study interval contains statistically significant
363 anticlustered arrangements. Presence of a statistically significant anticlustered arrangement on the
364 right side was not identified by the analysis of the average *K*-function. Second, the number of red
365 "pillars" or bars at each length scale determines the number of clusters. If we draw a horizontal
366 line passing through the length scale h ($y = h$) and then count the number of red pillars that it
367 intersects, this is the number of clusters. Accordingly, the width of each red pillar at every length
368 scale is the size of that cluster. Third, as length scale increases, the adjacent smaller clusters merge
369 and form larger clustered structures. Merging of clusters is manifested by "dome-shaped"
370 configurations, for example between 7 and 17 m where multiple clusters merge and form one major
371 cluster at length scales larger than 4 m. In addition, Fig. 11b shows several thin localized clusters
372 that span only a few length scales. At larger length scales, these isolated clusters are averaged by
373 nearby anticlustered fractures, resulting in local arrangements that are not statistically
374 distinguishable from random, and thus they disappear from the profile. On the other hand, some

of the clusters are so pronounced that not only do they span across all length scales, but they also affect the interpretation of the local arrangement of nearby fractures as length scale increases. One example is the cluster between 3.8 and 5.2 m along the study interval. The extension of this cluster at various length scales is highlighted by a red, "wing-shaped" pattern centered at 4.5 m in Fig. 10.

Figure 11: Results of local analysis of the *K*-function at Pedernales. (a) Profile of local *p*-values. (b) Statistically significant clustered (red) and anticlustered arrangements (purple). For fractures along scanline (*x*-axis), *p*-values are calculated at every length scale (*y*-axis) based on 99 bootstrap realizations.

In comparison, **Fig. 12** shows a profile of the local *K*-function analysis in a random arrangement. Unlike the Pedernales dataset, the local analysis of the *K*-function in a random arrangement indicates only sparse regions of fracture clustering and anticlustering. No major statistically significant pattern or structure appears in this profile. All the isolated clusters and anticlusters disappear as length scale increases.

Figure 12: Profile of statistically significant clustered (red) and anticlustered (purple) arrangements identified by local analysis of the *K*-function along study interval in a synthetic random arrangement. Unlike the arrangement at Pedernales, statistically significant arrangements here are minor, and they disappear at larger length scales.

Results indicate that the local analysis of the *K*-function is more informative than the overall analysis, and it provides more information about the location, frequency, and properties of clustered and anticlustered arrangements along the study interval.

Note that the methods developed here help in understanding of the distribution of fractures in the subsurface by quantifying spatial arrangements. Such spatial analyses are not necessarily sufficient to explain the genesis of the clustering and anticlustering in the subsurface. Additional information and studies are required for identifying geological processes that could influence fracture spatial arrangements, such as faulting, folding, diagenesis, and boudinage. The Pedernales outcrop is a

403 relatively undeformed panel of limestone with approximately constant thickness. Therefore, it
404 would be unlikely that any faulting, folding, or boudinage control the clustering of the veins.
405 Rather, fracture-fracture interaction during subcritical crack growth with modification of the local
406 stress field near fracture tips may have caused the observed clustering (c.f., Olson, 2004).
407 However, to reliably show any relationship (or lack thereof) between subsurface processes and the
408 clustering of the veins at Pedernales, a comprehensive modeling workflow is required.

409 # 6. Simulation of fracture arrangements

410 In subsurface studies, one may consider generating fracture model realizations based on 1D
411 fracture arrangements observed in the field from outcrop scanlines, borehole image logs, or cores.
412 For example, in modeling naturally fractured reservoirs, several discrete fracture network
413 realizations may be generated for uncertainty quantification, modeling hydrocarbon production
414 (e.g., Shakiba, 2014, 2018; Nejadi et al., 2017; Sun et al., 2017), or investigating hydraulic fracture
415 propagation (e.g., Olson, 2004; Wu and Olson, 2016; Weng et al., 2016). Therefore, it is important
416 for the generated fracture models to honor multiscale spatial properties and arrangement of the
417 fractures observed in the field.

418 ## 6.1 Simulation Technique

419 In this section, we develop a simulation technique to statistically reconstruct multiscale spatial
420 arrangement of fractures in model realizations. The simulation starts with a random fracture
421 arrangement. A fracture is randomly selected and relocated along the study interval (i.e., randomly
422 perturbed). The K-function is then calculated for the new arrangement and compared against the
423 K-function of the original dataset. The new arrangement is accepted, if the mismatch between the
424 K-functions is reduced (i.e., a greedy approach); otherwise, it is rejected. The simulation follows
425 an iterative approach whereby each successful iteration brings the candidate realization closer to

426 the target spatial arrangement. This simulation technique is formulated as a non-parametric
427 optimization problem. An algorithm based on random perturbation is deployed to find the solution
428 (Tscheschel and Stoyan, 2006; Illian et al., 2008; Wiegand and Moloney, 2013). A flow chart of
429 the proposed simulation technique is shown in **Fig. 13**.

430 The main advantage of this simulation technique is that it can incorporate any complex spatial
431 pattern as long as it can be quantified during optimization. Various summary statistics can be used
432 to calculate the error or the mismatch (loss function) between a simulated arrangement and the
433 target arrangement after each iteration. Here, loss function is defined as the L^2 norm of the error
434 in the *K*-function over various length scales to capture multiscale characteristics of fracture
435 arrangements. Therefore, the loss function E_j after iteration j is described by

$$E_j = \sqrt{\sum_{h=h_1}^{h_n} w_j \left[K_j^*(h) - K(h)\right]^2} \tag{9}$$

436 where n is the total number of length scales; $K_j^*(h)$ is the *K*-function at the length scale h calculated
437 for the realization in iteration j; $K(h)$ is the *K*-function at the length scale h calculated for the
438 observed fractures; and w_j is the weighting factor for each length scale (different from the
439 weighting factor ω_i for edge correction in Eq. 2). If simulation of spatial arrangement at a specific
440 length scale is of more interest, a greater value should be selected for the associated weighting
441 factor (w_j). Note that the simulation technique automatically honors fracture intensity. Each
442 fracture realization has the same number of fractures along the study interval as that of the original
443 dataset. Since fractures are only relocated during iterations, the overall fracture intensity remains
444 constant.

Figure 13: Flowchart describing the simulation technique for 1D fracture arrangements.

6.2 Simulation of Fracture Arrangement at Pedernales

Fig. 14 illustrates the simulation results of the Pedernales fractures. The 1D fracture arrangements generated after various iterations are shown in this figure. Based on visual inspection, the final simulated fracture arrangement (after 100,000 iterations) is similar to the target fracture arrangement at Pedernales. In this simulated arrangement, fractures are primarily clustered around one side of the study domain, and multiple smaller clusters can be observed.

Figure 14: Comparison of original fracture arrangement at Pedernales (top) with simulated fracture arrangements after various numbers of iterations.

In addition, good agreement exists between statistical properties of the final simulated arrangement and those of the Pedernales dataset (**Fig. 15**). Calculated values of K_n for the simulated arrangement after 100,000 iterations are similar to those of the original dataset at the pre-specified length scales of 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, and 10 m. The only discrepancy in the *K*-function values occurs at the smallest length scale (10^{-4} m). Often more iterations are needed to match spatial arrangement at small length scales. Comparison of the *K*-functions between candidate realizations and the Pedernales fractures (right column in Fig. 15) shows that initial iterations capture the spatial arrangement at larger scales. As the simulation proceeds, the spatial arrangements at small scales match as well. At small length scales, the fracture spatial arrangement is dominated by only a few clusters. To match these fine patterns, fractures should be randomly placed near each other, which often requires many more iterations.

Although the optimization algorithm is primarily formulated to minimize the mismatch in the *K*-function, a good match still exists between the histograms and the cumulative size distributions of

fracture spacing. This indicates that some information redundancy may exist between these summary statistics. Therefore, the *K*-function can capture multiscale spatial characteristics of fracture arrangement at Pedernales.

Figure 15: Comparison of histograms of spacing (left column), cumulative distribution functions of spacing (middle column), and normalized K-functions (right column) calculated for Pedernales dataset with those of simulated fracture arrangements after various iterations (rows). As number of iterations increases, deviations between statistical properties of simulated fracture arrangements and those of Pedernales fractures decrease.

Fig. 16 shows how the loss function (total error) reduces as the simulation proceeds. After 100,000 iterations, a satisfactorily small error is achieved. Although the number of iterations is large, the total computational time is reasonable. Since only one fracture is randomly relocated in each simulation step, the spatial arrangement remains mostly unchanged from one iteration to another. Storing and recycling variables and calculations from previous iterations greatly reduce the computational time. Moreover, selecting proper stopping criteria (such as desired minimum error, maximum number of iterations, maximum number of iterations with no error improvement, etc.) prevents the optimization algorithm from conducting unnecessary iterations.

Figure 16: Loss function (total error) after each iteration calculated using Eq. 9.

6.3 Filling Measurement Gaps

In addition to the simulation of fracture arrangements, the proposed technique can be adopted for restoring data gaps in measurements. Not all fractures are observed during a field measurement; for example, an outcrop might be covered by vegetation, soil, or water (e.g., Peacock et al., 2019). Or a core might be greatly damaged (e.g., Li et al., 2018). Such gaps can introduce error into the spatial analysis of fracture arrangements. To restore fracture arrangement across the entire study interval, the simulation technique can be deployed to generate a realization of fracture locations in

the measurement gaps, a process known as statistical imputation. The synthetic example below demonstrates the statistical imputation using the simulation technique.

Here we created a 60-m-long synthetic fracture arrangement with a gap in the middle. The two sections with available fracture locations and the gap are each 20 m long. Fracture locations are generated by superimposing a Thomas cluster point pattern on a random arrangement with intensity of 10 fractures/m (**Fig. 17**). A Thomas point pattern belongs to the family of Neyman-Scott point processes (Neyman and Scott, 1952) wherein clusters are randomly located along the study interval. The central points of clusters are called "parents," and the subsequent points inside the clusters are called "daughters." Distances between daughters and the corresponding parents follow a zero-mean normal Gaussian distribution with the variance σ^2. Here three fracture clusters are randomly placed on each side of the study interval. Within each cluster, there are 100 daughter fractures. Distances between daughters and parents in each cluster follow a normal Gaussian distribution with the standard deviation σ of 0.8 m.

Figure 17: A synthetic fracture arrangement with a gap in the middle. Fracture arrangement is created by superimposing a Thomas cluster point pattern on a random arrangement with intensity of 10 fractures/m.

To restore the gap in the middle, the simulation technique needs a reference spatial arrangement for the optimization problem (i.e., reference values of the *K*-function at multiple scales). Our assumption here is that the fracture arrangement in the middle is similar to that in the rest of the study interval. Since the fractures on both sides of the study interval are positioned using the same approach, either side (or the average of the two) can be selected as the reference. In this work, we chose the right side as the reference during optimization.

Figs. 18 and **19** depict the restored fracture arrangement and the corresponding *K*-functions for both sides of the study interval as well as for the simulated section in the middle. The simulated fracture arrangement in the middle visually resembles the arrangement in the rest of the study interval. Two or three clusters can be identified in the middle section, and fracture spacings inside these clusters approximately follow a Gaussian distribution. Moreover, comparison of the *K*-functions calculated for each section indicates that all three sections have similar spatial arrangements over various length scales.

Figure 18: Restored synthetic arrangement. The gap in fracture arrangement, in the middle, is restored using the simulation technique.

Figure 19: Comparison of the *K*-functions calculated for the left, middle, and right sections along the study interval. Simulated fractures in the middle exhibit a spatial arrangement similar to those of the left and right sections.

Replacing the data gap with a spatially similar arrangement allows a thorough evaluation of fracture arrangement along the study interval. Now, with the middle section restored, formerly segregated fractures can be pooled, and the entire length of the study interval can be studied as a whole (**Fig. 20**). If the gap is not restored, it would erroneously inflate the degree of clustering. For example, with the gap, the spatial arrangement is significantly clustered between 2×10^{-4} and 2×10^{-3} m length scales. However, after restoring the gap, the spatial arrangement becomes indistinguishable from random at those length scales. Similarly, the spatial arrangement changes from clustered to indistinguishable from random at length scales larger than 3 m after restoring the gap. Therefore, missing data imputation is crucial when the goal is to analyze the entire length of the study interval.

Figure 20: Comparison of the *K*-functions calculated for the synthetic fracture arrangement before and after restoring the gap (i.e., data imputation). Spatial analysis of fracture arrangement with the gap erroneously shows a higher degree of clustering.

One potential application of our simulation technique is to predict fracture arrangements beyond the boundaries of the study domain and at larger length scales than the scale of measurement. As shown in Fig. 19, spatial analysis of the fracture arrangement was initially limited to length scales of less than 10 m (half-length of the sections with available fracture data) because of the gap. However, after data imputation, we extended the spatial analysis to a 30-m length scale (half-length of the whole study interval). Therefore, using the simulation technique fracture measurements made on the scale of an outcrop or a borehole image log, for example, can be extended to subsurface reservoir scales for further modeling. Or, fractures measured in a core can be used to generate synthetic fracture locations along the whole trajectory of the associated well drilled through a formation of interest, while statistically preserving the spatial arrangement of fractures. However, note that prediction of fracture arrangement beyond a visible sampling domain and over larger length scales could be risky compared to restoration of measurement gaps inside the boundaries of a study domain and at the observed range of length scales. Fracture arrangements may greatly change from one region to another because of anisotropy and heterogeneity in subsurface rock properties and stress states. For example, the left half of the scanline measurement at Pedernales has a different arrangement from that of the right half. Prediction of fracture arrangement in one half based on the observed fracture arrangement in the other half would produce erroneous results. Therefore, prediction of fracture arrangements beyond the boundaries of study domains must be corroborated by other relevant studies and field observations.

Conclusions

Ripley's *K*-function quantifies spatial arrangement of fractures in one-dimensional measurements to identify statistically significant clustered and anticlustered arrangements. The average value of the *K*-function is used to evaluate the overall fracture clustering and anticlustering, whereas local values of the *K*-function are used in understanding the changing nature of fracture arrangement as a function of position along the study interval. Local analysis is more informative than the overall analysis of spatial arrangement, because it indicates where clustered and anticlustered arrangements occur along the study interval.

In addition to spatial characterizations, the *K*-function can be deployed for simulation of fracture arrangements. Practitioners can use the simulation technique to statistically reconstruct fracture arrangements that mimic spatial arrangements observed in the field. This is a powerful capability that expands and improve our ability to analyze and model naturally occurring fracture arrangements. Our simulation technique can be also used for missing data imputation to restore spatial gaps in measurements, such as in partially exposed outcrops, or for quality control to validate data of low resolution or distorted measurements, such as in borehole image logs.

Powerful yet simple formulation of Ripley's *K*-function, as well as simulation and data imputation techniques, can be readily applied to various 1D fracture datasets from scanline measurements, borehole image logs, and subsurface cores. Note that the methods we presented here primarily apply to fractures from the same set. Fractures can be classified into various sets/groups based on orientation, size, aperture filling condition, and other attributes. When fractures from multiple sets exist in a dataset, the spatial relationships between different sets become as important as the spatial arrangements within each set. Methods based on marked (labeled) point processes are good candidates for such spatial analyses.

Moreover, Ripley's *K*-function can be extended to 2D and 3D applications. One approach is to replace 1D search intervals with circular or spherical windows for 2D and 3D applications, respectively. Consequently, new weighting factors should be derived to correct for edge effects in higher dimensions. The challenge, however, will be how to transform fractures into discrete points in higher dimensions. One solution would be to use barycenters of fracture traces or intersection points. Such potential extensions of the *K*-function analysis should be tested by field data measurements.

Acknowledgments

This research was funded by grant DE-FG02-03ER15430 from Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy, and by the Fracture Research and Application Consortium. We thank Randy Marrett for providing feedback on an early version of this work. We also thank him, Leonel Gomez, Celinda Kelsey, and Chris Hare for collecting data at Pedernales. We are grateful to Steve Laubach, William Dunne, Mark Fischer, and Paul Gillespie for their thoughtful review comments on this work.

References

Baddeley, A., Bárány, I., and Schneider, R., 2007. Spatial point processes and their applications. Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 1–75.

Bagni, F.L., Bezerra, F.H., Balsamo, F., Maia, R. P., and Dall'Aglio, M., 2020. Karst dissolution along fracture corridors in an anticline hinge, Jandaíra Formation, Brazil: Implications for reservoir quality. Marine and Petroleum Geology, 115, 104249.

Berkowitz, B., 2002. Characterizing flow and transport in fractured geological media: a review. Advances in Water Resources, 25, 861–884.

Bistacchi, A., Mittempergher, S., Martinelli, M., and Storti, F., 2020. On a new robust workflow for the statistical and spatial analysis of fracture data collected with scanlines (or the importance of stationarity). Solid Earth, 11(6), 2535–2547.

Borgos, H.G., 2000. Partitioning of a line segment. Stochastic Modeling and Statistical Inference of Geological Fault Populations and Patterns. Norwegian University of Science and Technology.

Cressie, N., 2015. Statistics for spatial data. John Wiley & Sons Inc.

Dichiarante, A.M., McCaffrey, K.J., Holdsworth, R.E., Bjørnarå, T.I., and Dempsey, E.D., 2020. Fracture attribute scaling and connectivity in the Devonian Orcadian Basin with implications for geologically equivalent sub-surface fractured reservoirs. Solid Earth, 11(6), 2221–2244.

Dixon, P.M., 2014. Ripley's K Function. Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat07751

Du Bernard, X., Labaume, P., Darcel, C., Davy, P., and Bour, O., 2002. Cataclastic slip band distribution in normal fault damage zones, Nubian sandstones, Suez rift. Journal of Geophysical Research, 107 (B7), ETG-6.

Efron, B., 1992. Bootstrap methods: another look at the jackknife. Breakthroughs in Statistics. Springer, 569–593.

Erlich, R.N., and Coleman, J.L., 2005. Drowning of the Upper Marble Falls carbonate platform (Pennsylvanian), central Texas: A case of conflicting signals? Sedimentary Geology, 175, 479–499.

Gale, J.F.W., Laubach, S.E., Olson, J.E., Eichhubl, P., and Fall, A. 2014. Natural fractures in shale: a review and new observations. AAPG Bulletin, 98(11), 2165–2216.

Gauthier, B.D.M., Auzias, V., Garcia, M., and Chiapello, E., 2002. Static and dynamic characterization of fracture pattern in the Upper Jurassic reservoirs of an offshore Abu Dhabi field: From well data to full field modeling. SPE Abu Dhabi International Petroleum Exhibition and Conference. Abu Dhabi, United Arab Emirates.

Gillespie, P.A., Howard, C.B., Walsh, J.J., and Watterson, J., 1993. Measurement and characterisation of spatial distributions of fractures. Tectonophysics, 226(1–4), 113–141.

Gillespie, P.A., Johnston, J.D., Loriga, M.A., McCaffrey, K.J.W., Walsh, J.J., and Watterson, J., 1999. Influence of layering on vein systematics in line samples. Geological Society, London, Special Publications 155(1), 35–56.

Gillespie, P.A., Walsh, J.J., Watterson, J., Bonson, C.G., and Manzocchi, T., 2001. Scaling relationships of joint and vein arrays from The Burren, Co. Clare, Ireland. Journal of Structural Geology, 23(2-3), 183–201.

Gomez, L.A., 2007. Characterization of the Spatial Arrangement of Opening-Mode Fractures. PhD dissertation, The University of Texas at Austin.

649 Grassberger, P., and Procaccia, I., 2004. Measuring the strangeness of strange attractors. The
650 Theory of Chaotic Attractors, Springer, 170–189.

651 Hanke, J.R., Fischer, M.P., and Pollyea, R.M., 2018. Directional semivariogram analysis to
652 identify and rank controls on the spatial variability of fracture networks. Journal of Structural
653 Geology, 108, 34–51.

654 Hardebol, N. J., Maier, C., Nick, H., Geiger, S., Bertotti, G., and Boro, H. 2015. Multiscale fracture
655 network characterization and impact on flow: A case study on the Latemar carbonate platform,
656 Journal of Geophysical Research: Solid Earth, 120, 8197– 8222.

657 Hesterberg, T., 2011. Bootstrap. Wiley Interdisciplinary Reviews: Computational Statistics, 3(6),
658 497–526.

659 Hooker, J.N., Gale, J.F.W., Gomez, L.A., Laubach, S.E., Marrett, R., and Reed, R.M., 2009.
660 Aperture-size scaling variations in a low-strain opening-mode fracture set, Cozzette Sandstone,
661 Colorado. Journal of Structural Geology, 31(7), 707–718.

662 Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D., 2008. Statistical analysis and modelling of
663 spatial point patterns. John Wiley & Sons Inc.

664 Laubach, S.E., 1992. Attributes of fracture networks in selected Cretaceous sandstones of the
665 Green River and San Juan basins. Geological Studies Relevant to Horizontal Drilling: Examples
666 from Western North America. Denver, Colorado, Rocky Mountain Association of Geologists, 61–
667 74.

668 Laubach, S.E., Mace, R.E., Nance, H.S., and Rossmanith, H.P., 1995. Fault and joint swarms in a
669 normal fault zone. *In:* Rossmanith, H.-P., (Ed.), 1995. Mechanics of jointed and faulted rock.
670 Balkema, Rotterdam, 305–309.

671 Laubach, S.E., 2003. Practical approaches to identifying sealed and open fractures. AAPG
672 Bulletin, 87(4), 561–579.

673 Laubach, S.E., Lamarche, J., Gauthier, B.D., Dunne, W.M., and Sanderson, D.J., 2018. Spatial
674 arrangement of faults and opening-mode fractures. Journal of Structural Geology, 108, 2–15.

675 Laubach, S.E., Lander, R.H., Criscenti, L.J., et al., 2019. The role of chemistry in fracture pattern
676 development and opportunities to advance interpretations of geological materials. Reviews of
677 Geophysics, 57 (3), 1065–1111.

678 Li, J.Z., Laubach, S.E., Gale, J.F.W., and Marrett, R., 2018. Quantifying opening-mode fracture
679 spatial organization in horizontal wellbore image logs, core and outcrop: Application to Upper
680 Cretaceous Frontier Formation tight gas sandstones, USA. Journal of Structural Geology, 108,
681 137–156.

682 Marcon, E., and Puech, F., 2003. Evaluating the geographic concentration of industries using
683 distance-based methods, Journal of Economic Geography, 3(4), 409–428,

Marrett, R., Gale, J.F., Gómez, L.A., and Laubach, S.E., 2018. Correlation analysis of fracture arrangement in space. Journal of Structural Geology, 108, 16–33.

Marrett, R., Ortega, O.J., and Kelsey, C.M., 1999. Extent of power-law scaling for natural fractures in rock. Geology 27(9), 799–802.

Narr, W., Schechter, D.S., and Thompson, L.B., 2006. Naturally fractured reservoir characterization. Society of Petroleum Engineers, Richardson, TX

Narr, W., and Suppe, J., 1991. Joint spacing in sedimentary rocks. Journal of Structural Geology, 13, 1037–1048.

Nejadi, S., Trivedi, J.J., and Leung, J., 2017. History matching and uncertainty quantification of discrete fracture network models in fractured reservoirs. Journal of Petroleum Science and Engineering, 152, 21–32.

Neyman, J., and Scott, E.L., 1952. A theory of the spatial distribution of galaxies. Astrophysical Journal, 116, 144–163.

Odling, N.E., Gillespie, P., Bourgine, B., Castaing, C., Chiles, J.P., Christensen, N.P., Fillion, E., Genter, A., Olsen, C., Thrane, L., and Trice, R., 1999. Variations in fracture system geometry and their implications for fluid flow in fractures hydrocarbon reservoirs. Petroleum Geoscience, 5(4), 373–384.

Olson, J.E., 2004. Predicting fracture swarms—The influence of subcritical crack growth and the crack-tip process zone on joint spacing in rock. *In:* Cosgrove, J.W., and Engelder, T., (eds) 2004. The Initiation, Propagation, and Arrest of Joints and Other Fractures. Geological Society, London, Special Publications, 231, 73–88.

Ortega, O.J., Marrett, R., and Laubach, S.E., 2006. A scale-independent approach to fracture intensity and average fracture spacing: AAPG Bulletin, 90(2), 193–208.

Peacock, D.C.P., Sanderson, D.J., Bastesen, E., Rotevatn, A., and Storstein, T.H., 2019. Causes of bias and uncertainty in fracture network analysis. Norwegian Journal of Geology, 99, 1–16.

Pyrcz, M.J., and Deutsch, C.V., 2014. Geostatistical reservoir modeling. Oxford University Press.

Questiaux, J.M., Couples, G.D., and Ruby, N., 2010. Fractured reservoirs with fracture corridors. Geophysical Prospecting, 58(2), 279–295.

Ripley, B.D., 1976. The second-order analysis of stationary point processes. Journal of Applied Probability, 13, 255–266.

Ripley, B.D. 1977. Modelling spatial patterns. Journal of the Royal Statistical Society, Series B, 39, 172–192.

Roy, A., Perfect, E., Dunne, W.M., and McKay, L.D., 2014. A technique for revealing scale-dependent patterns in fracture spacing data. Journal of Geophysical Research: Solid Earth, 119(7), 5979–5986.

719 Sanderson, D.J., and Peacock, D.C., 2019. Line sampling of fracture swarms and corridors. Journal
720 of Structural Geology, 122, 27–37.

721 Shakiba, M., 2014. Modeling and simulation of fluid flow in naturally and hydraulically fractured
722 reservoirs using embedded discrete fracture model (EDFM). MS Thesis, The University of Texas
723 at Austin.

724 Shakiba, M., de Araujo Cavalcante Filho, J.S., and Sepehrnoori, K., 2018. Using embedded
725 discrete fracture model (EDFM) in numerical simulation of complex hydraulic fracture networks
726 calibrated by microseismic monitoring data. Journal of Natural Gas Science and Engineering, 55,
727 495–507.

728 Smith, L., and Schwartz, F.W., 1984. An analysis of the influence of fracture geometry on mass
729 transport in fractured media. Water Resources Research, 20(9), 1241–52.

730 Smith, T.E., 2004. A scale-insensitive test of attraction and repulsion between spatial points
731 patterns. Geographical Analysis, 36, 315–331.

732 Stein, M.L., Quashnock, J.M., and Loh, J.M., 2000. Estimating the K function of a point process
733 with an application to cosmology. The Annals of Statistics, 28(6), 1503–1532.

734 Stoyan, D., and Penttinen, A. 2000. Recent applications of point process methods in forestry
735 statistics. Statistical Science, 15, 61–78.

736 Sun, W., Hui, M.H., and Durlofsky, L.J., 2017. Production forecasting and uncertainty
737 quantification for naturally fractured reservoirs using a new data-space inversion procedure.
738 Computational Geosciences, 21(5), 1443–1458.

739 Tscheschel, A., and Stoyan, D., 2006. Statistical reconstruction of random point patterns.
740 Computational Statistics & Data Analysis, 51(2), 859–871.

741 Wang, Q., Laubach, S.E., Gale, J.F.W., and Ramos, M.J., 2019. Quantified fracture (joint)
742 clustering in Archean basement, Wyoming: application of the normalized correlation count
743 method. Petroleum Geosciences, 25(4), 415–428.

744 Weng, X., Cohen, C.E., and Kresse, O., 2016. Impact of preexisting natural fractures on hydraulic
745 fracture simulation. Unconventional Oil and Gas Resources Handbook, Gulf Professional
746 Publishing, 289–331.

747 Wiegand, T., and Moloney, K.A., 2013. Handbook of spatial point-pattern analysis in ecology.
748 CRC Press.

749 Wright, W.R., Loucks, R.G., Gale, J.F.W., Kane, J., and McDonnell, A., 2007. Paleozoic reservoir
750 systems: Texas Hill Country-stratigraphy to petrophysics: a field trip for the Annual Meeting of
751 the Petrophysicists and Well Log Analysts, Guidebook, The University of Texas at Austin, Bureau
752 of Economic Geology, 77 p.

753 Wu, K., and Olson, J.E., 2016. Numerical investigation of complex hydraulic-fracture
754 development in naturally fractured reservoirs. SPE Production & Operations, 31(4), 300–309.

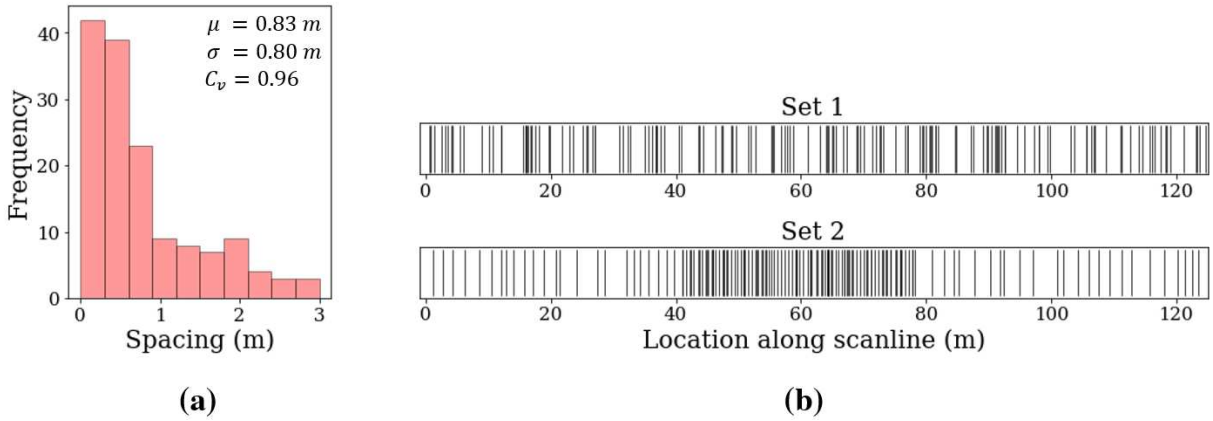
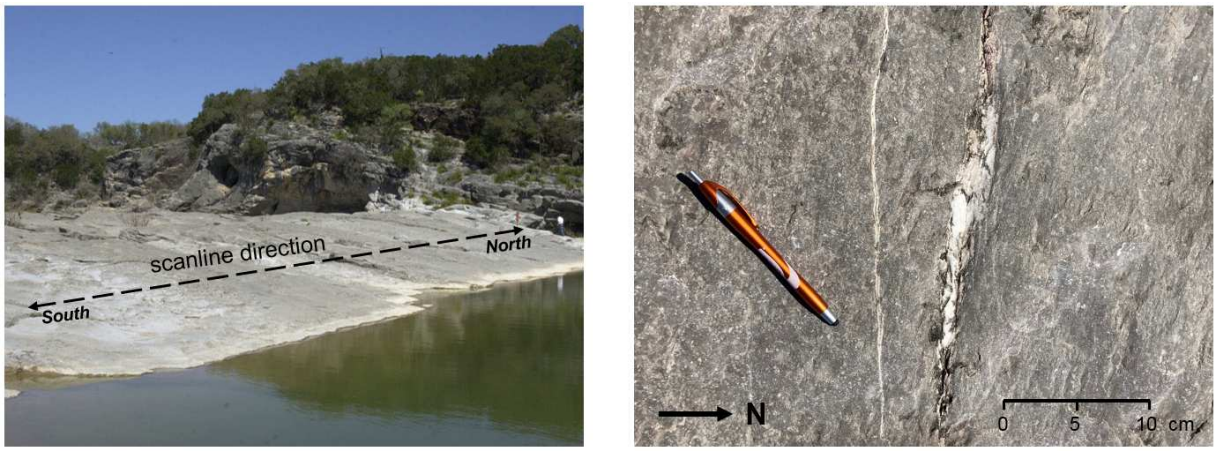


Figure 1: Comparison of spatial arrangements in two fracture sets. (a) Both fracture sets have identical mean μ, standard deviation σ, coefficient of variation C_V, and histograms of fracture spacing. (b) However, they exhibit different spatial arrangements. Set 2 is clustered, whereas set 1 is indistinguishable from a random arrangement.



(a) (b)

Figure 2: **Marble Falls Limestone and veins at Pedernales.** **(a) Limestone outcrop exposed on the bed of the Pedernales River in Pedernales Falls State Park, Texas. (b) Image of the calcite-filled veins.**

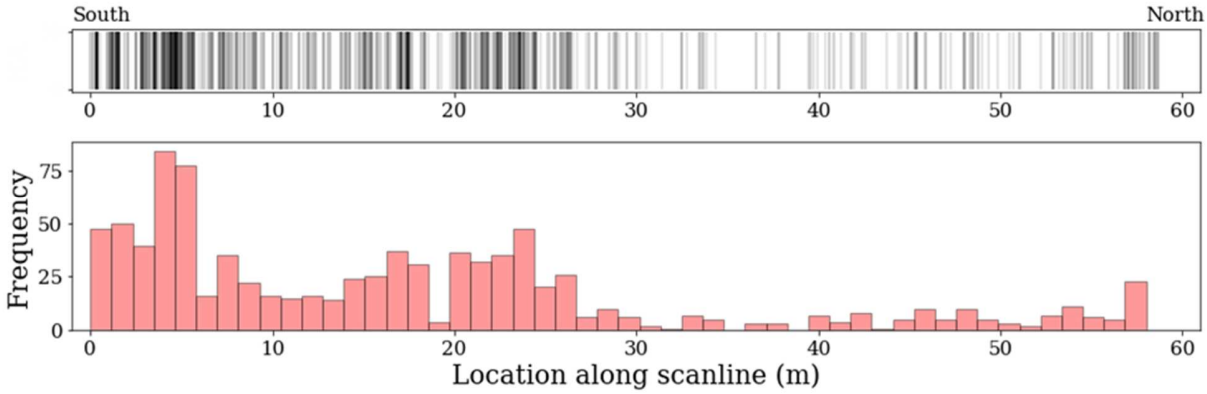


Figure 3: Visualization of fracture measurement at Pedernales. Stick plot indicating fracture locations (top), and variation of frequency over 50 equal-size bins of 1.16 m along the study interval (bottom).

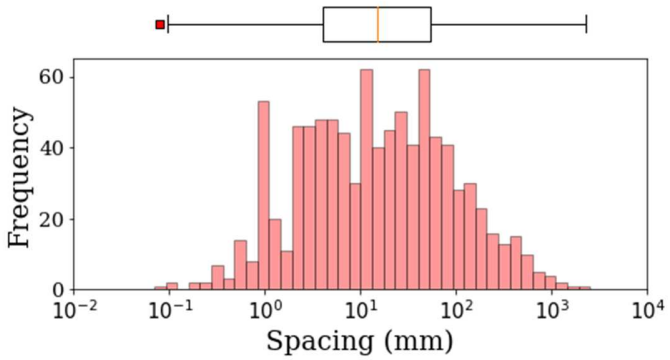


Figure 4: Histogram (bottom) and box plot (top) of fracture spacing at Pedernales. Spacings are on a logarithmic axis. Box plot displays the first and third quartiles (box edges), the median (middle line inside the box), variability outside the first and third quartiles (whiskers), and outliers (red squares) of spacing data.

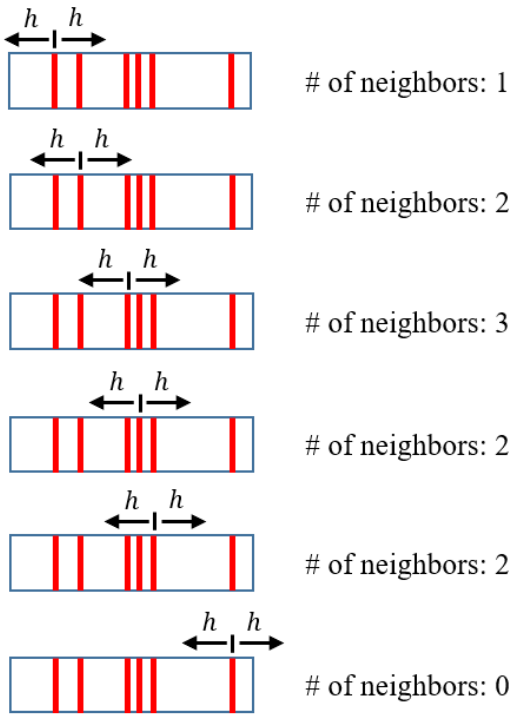


Figure 5: Example of fracture counting used in *K*-function calculation.

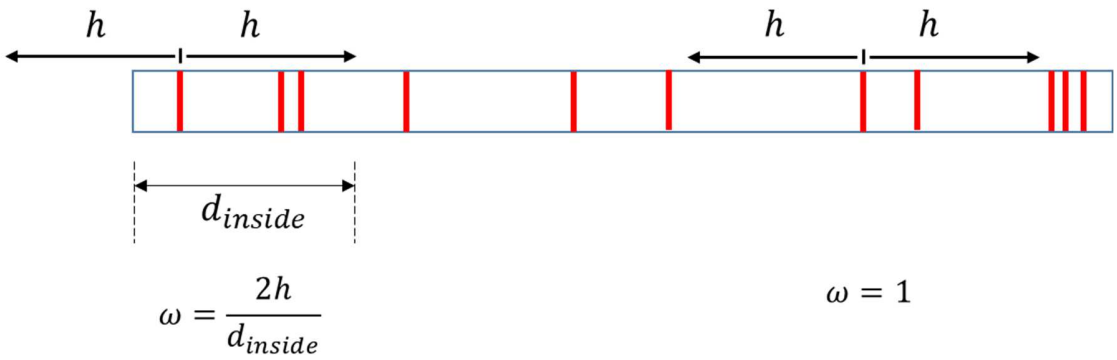


Figure 6: Illustration of edge effect. Part of the search interval shown on the left falls outside the study domain. Eq. 3 can be used to calculate weighting factor for edge correction. The search interval on the right, however, is completely inside the study domain; therefore, weighting factor is 1.

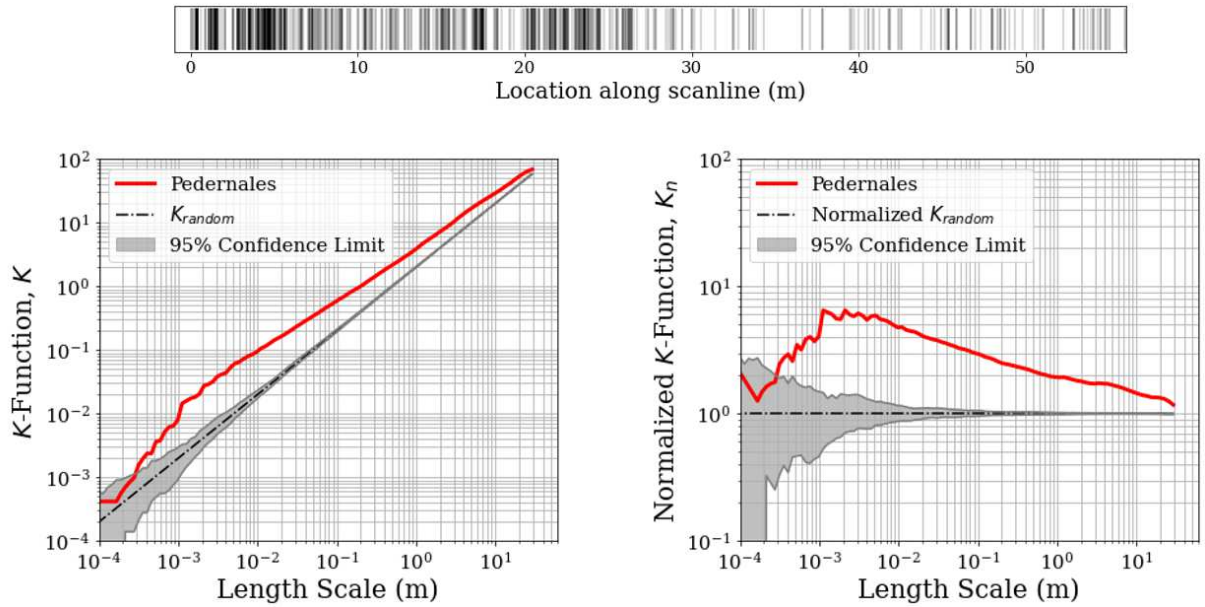


Figure 7: Spatial analysis of fracture arrangement at Pedernales. Stick plot showing fracture locations along scanline (top); calculated *K*-function (left); and normalized *K*-function (right). At length scales of less than 3×10^{-4} m, the fracture arrangement is indistinguishable from random, whereas at larger length scales a statistically significant clustered arrangement appears.

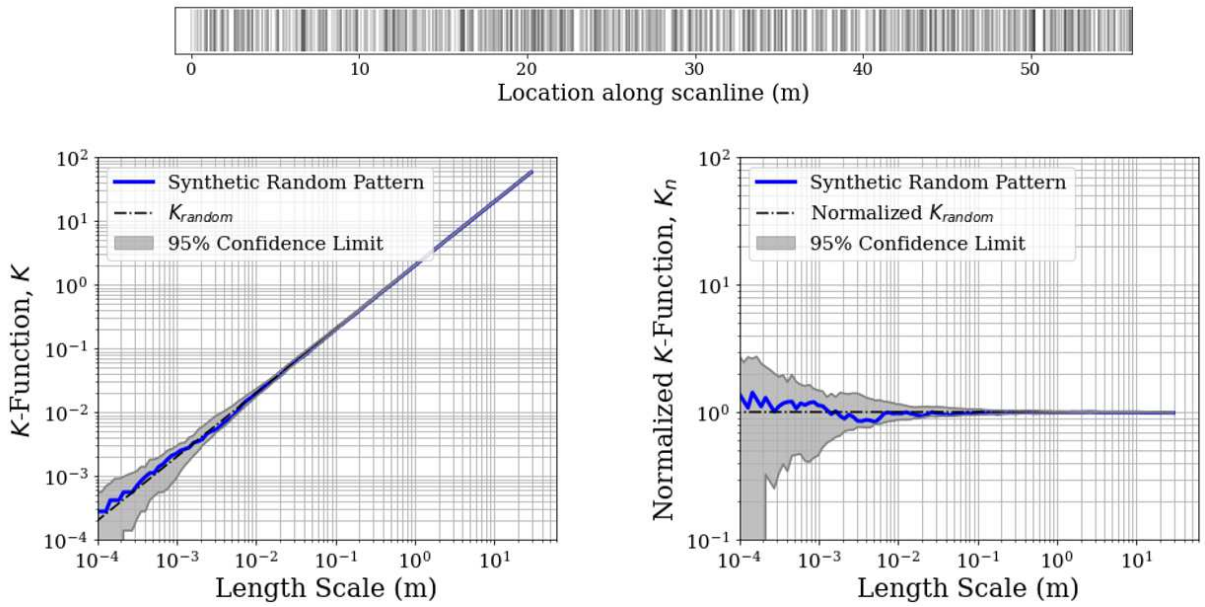


Figure 8: Spatial analysis of fracture arrangement in a dataset generated by random placement of fractures. Stick plot showing fracture locations along the study interval (top); calculated *K*-function (left); and normalized *K*-function (right). The *K*-function and normalized *K*-function fall inside confidence interval across all length scales.

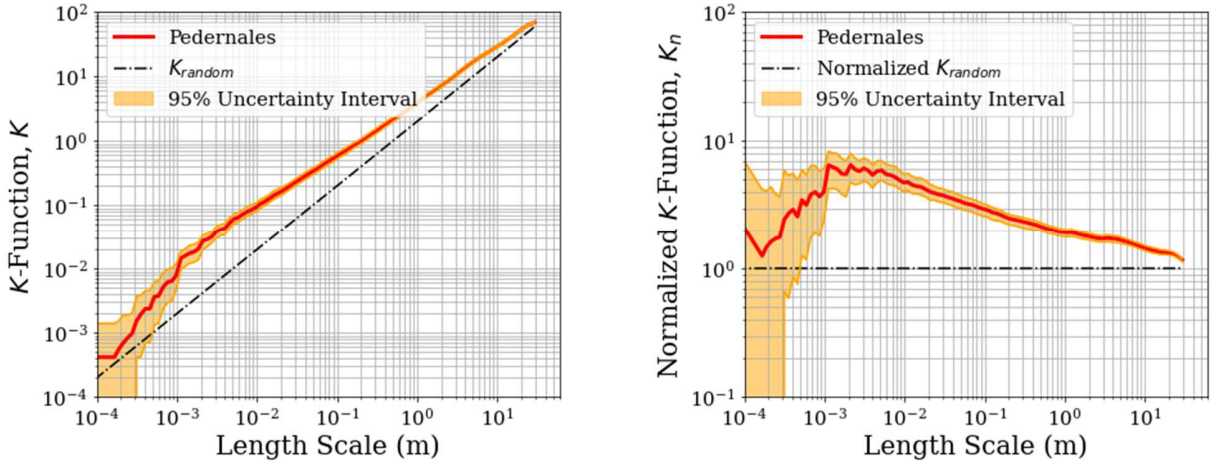


Figure 9: **Uncertainty intervals in the *K*-function and normalized *K*-function calculated for Pedernales dataset using statistical bootstrap.**

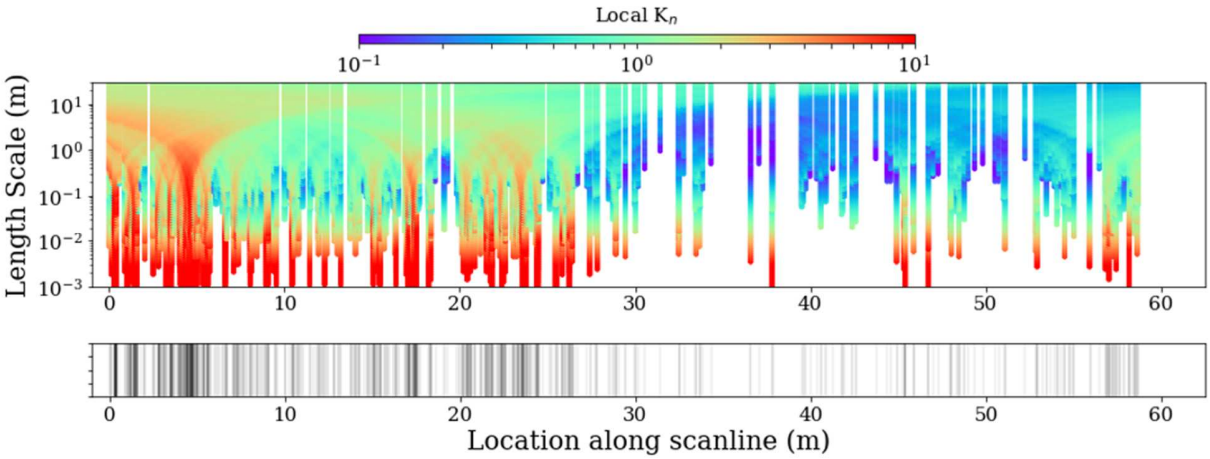


Figure 10: Profile of normalized local *K*-functions calculated for Pedernales fractures. For each fracture along scanline (*x*-axis), normalized local *K*-function is calculated at each length scale (*y*-axis). Points in the profile are colored with the associated values of normalized local *K*-function. Warm colors (values greater than 1) indicate fracture clustering, whereas cool colors (values less than 1) indicate fracture anticlustering.

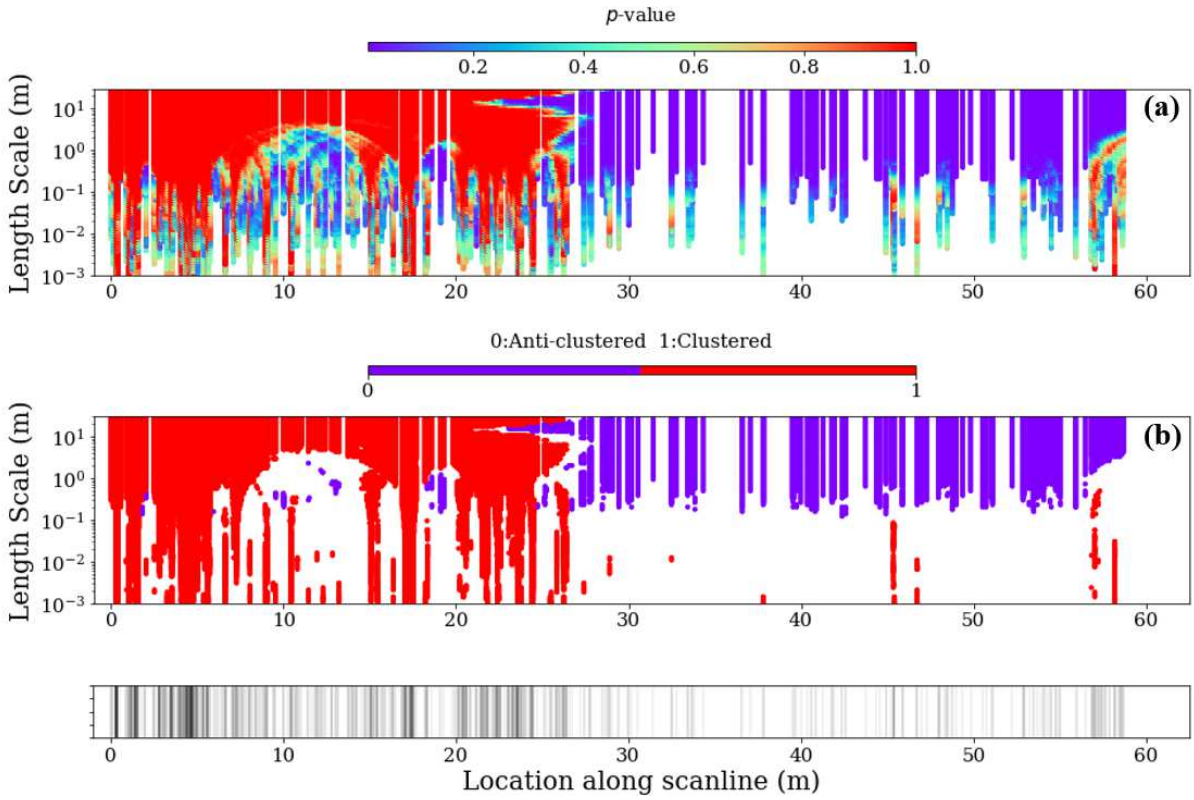


Figure 11: Results of local analysis of the *K*-function at Pedernales. (a) Profile of local *p*-values. (b) Statistically significant clustered (red) and anticlustered arrangements (purple). For fractures along scanline (*x*-axis), *p*-values are calculated at every length scale (*y*-axis) based on 99 bootstrap realizations.

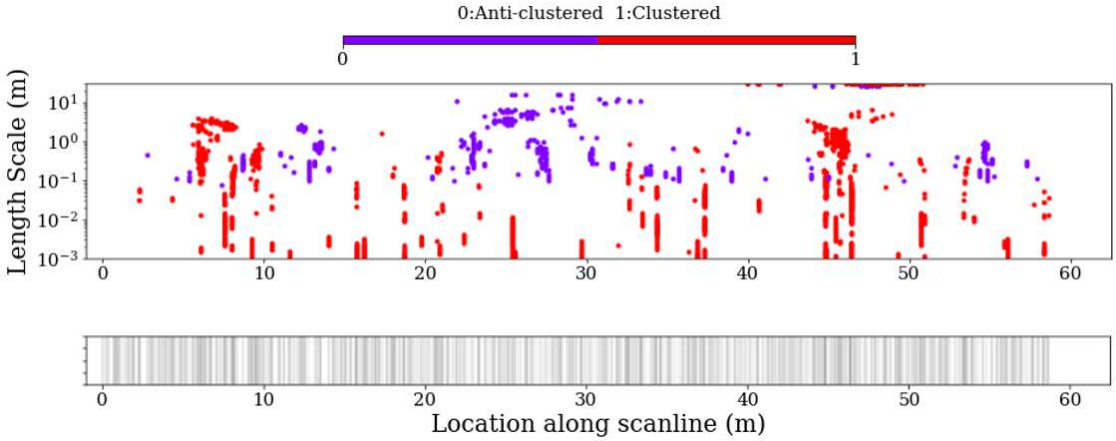


Figure 12: Profile of statistically significant clustered (red) and anticlustered (purple) arrangements identified by local analysis of the *K*-function along study interval in a synthetic random arrangement. Unlike the arrangement at Pedernales, statistically significant arrangements here are minor, and they disappear at larger length scales.

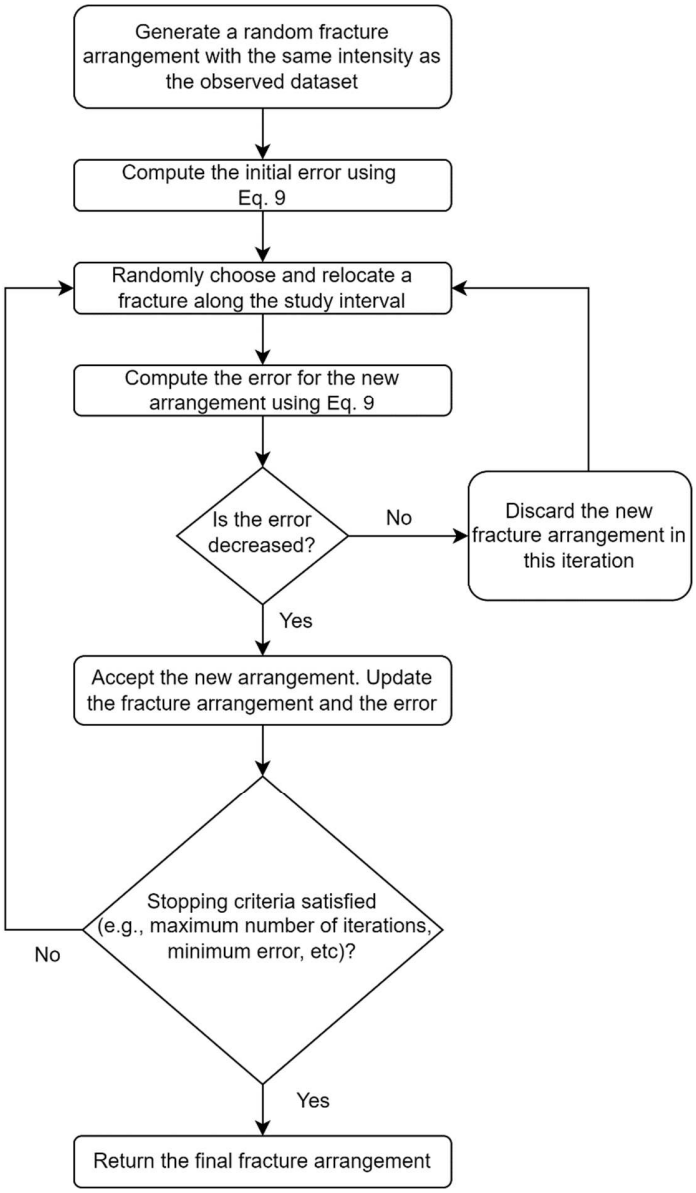


Figure 13: Flowchart describing the simulation technique for 1D fracture arrangements.

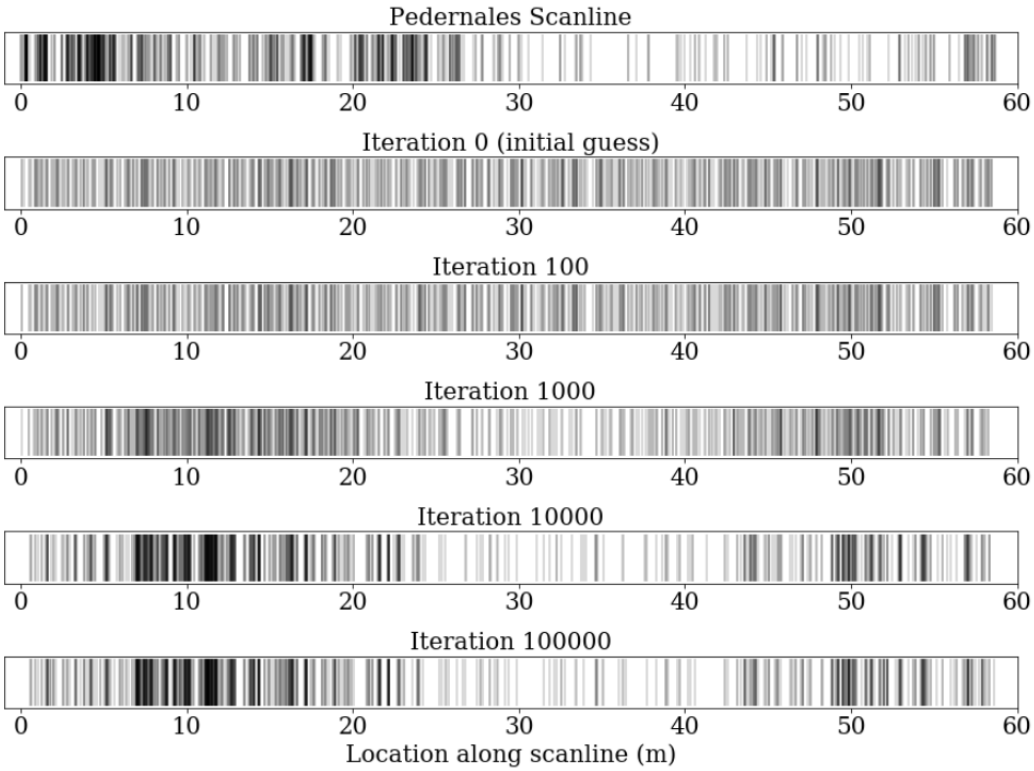


Figure 14: Comparison of original fracture arrangement at Pedernales (top) with simulated fracture arrangements after various numbers of iterations.

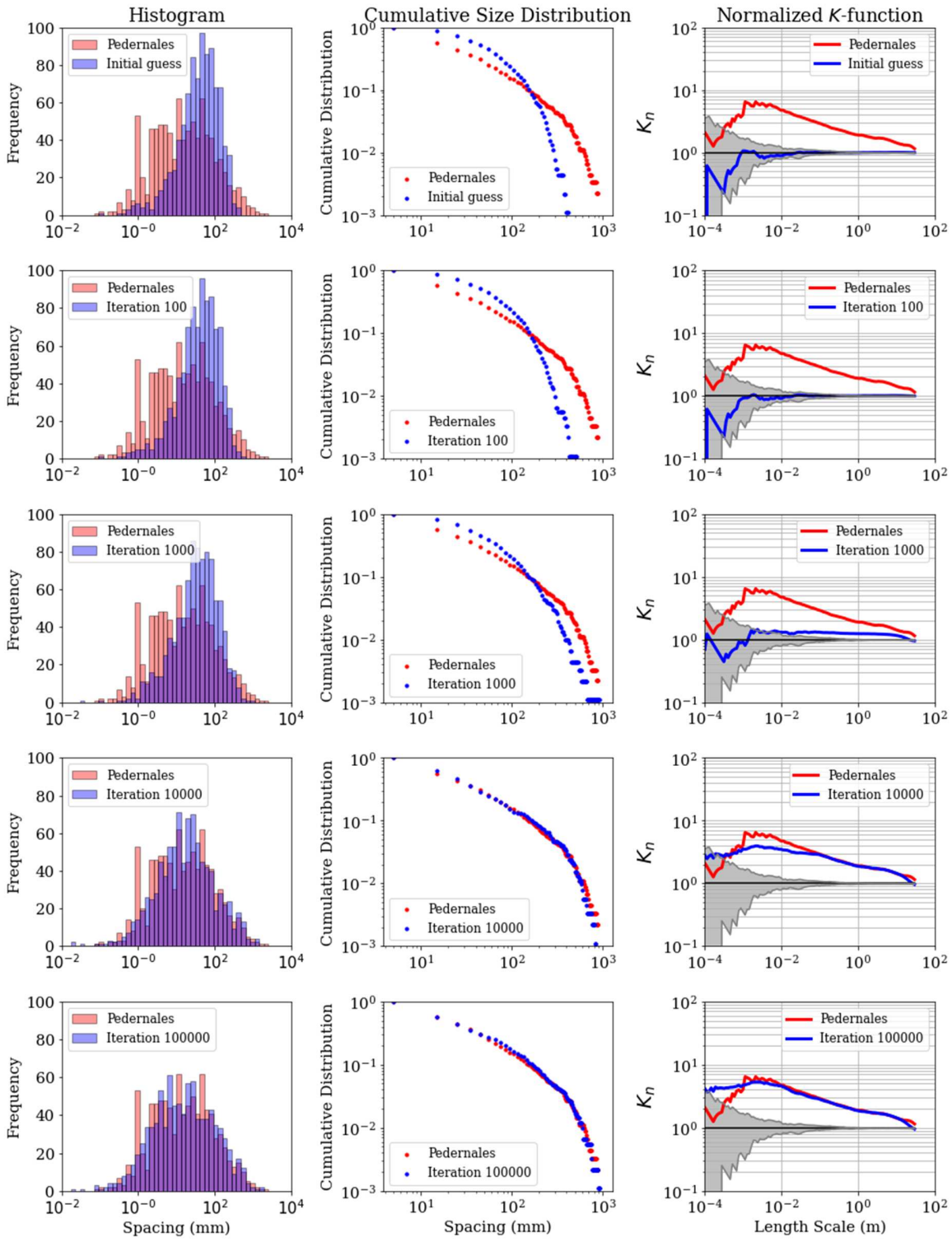


Figure 15: Comparison of histograms of spacing (left column), cumulative distribution functions of spacing (middle column), and normalized K-functions (right column) calculated for Pedernales dataset with those of simulated fracture arrangements after various iterations (rows). As number of iterations increases, deviations between statistical properties of simulated fracture arrangements and those of Pedernales fractures decrease.

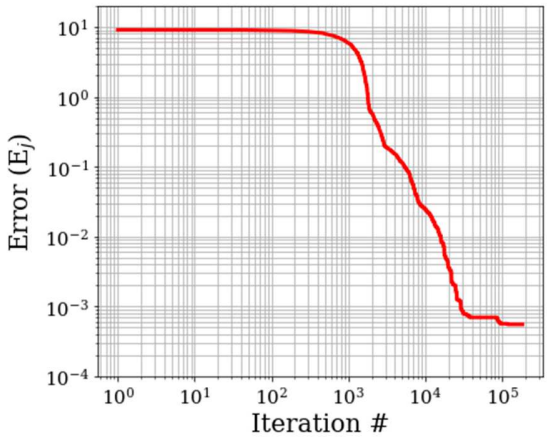


Figure 16: Loss function (total error) after each iteration calculated using Eq. 9.

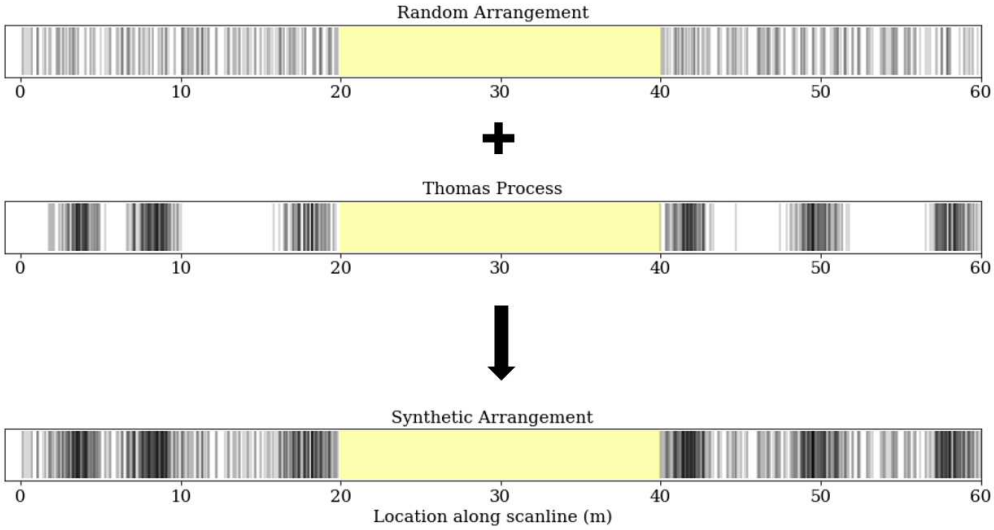


Figure 17: A synthetic fracture arrangement with a gap in the middle. Fracture arrangement is created by superimposing a Thomas cluster point pattern on a random arrangement with intensity of 10 fractures/m.

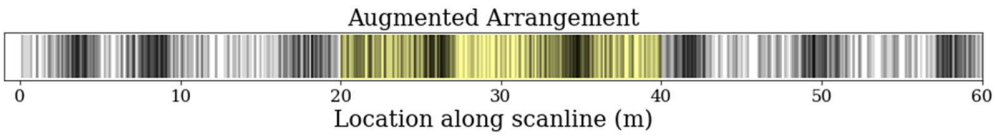


Figure 18: Restored synthetic arrangement. The gap in fracture arrangement, in the middle, is restored using the simulation technique.

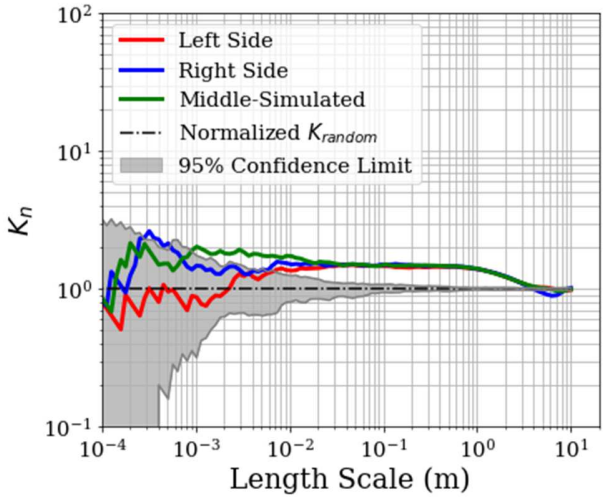


Figure 19: Comparison of the *K*-functions calculated for the left, middle, and right sections along the study interval. Simulated fractures in the middle exhibit a spatial arrangement similar to those of the left and right sections.

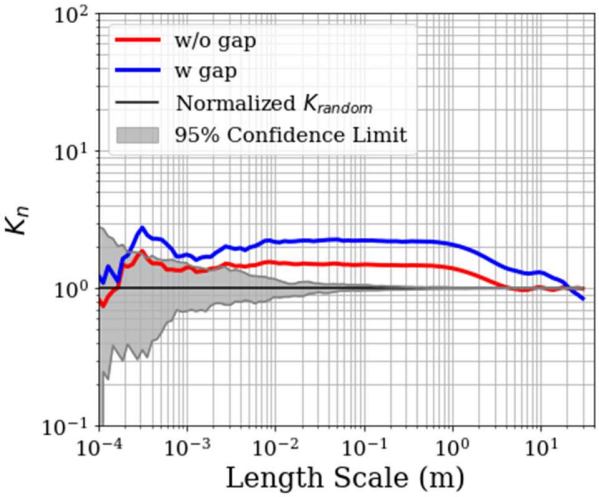


Figure 20: Comparison of the *K*-functions calculated for the synthetic fracture arrangement before and after restoring the gap (i.e., data imputation). Spatial analysis of fracture arrangement with the gap erroneously shows a higher degree of clustering.

	mean	std	min	25%	50%	75%	max
Spacing	64	155	0.08	4	15	53	2260

Table 1:Summary statistics for size distribution of fracture spacings at Pedernales. Units are millimeters.