

Highlights

FracG - An automated framework for discontinuity network characterization

Ulrich Kelka, Stefan Westerlund, Luk Peeters

- Automatic discontinuity network characterization
- Statistical description of fault and fracture networks
- Assessment of fluid flow direction

FracG - An automated framework for discontinuity network characterization^{★,★★}

Ulrich Kelka^{a,*}, Stefan Westerlund^a and Luk Peeters^b

^a

^b

^b

ARTICLE INFO

Keywords:

fault network
fracture network
statistical description
network analysis

ABSTRACT

This template helps you to create a properly formatted L^AT_EX manuscript.
\begin{abstract} ... \end{abstract} and \begin{keyword} ... \end{keyword} which contain the abstract and keywords respectively. Each keyword shall be separated by a \sep command.

1. Introduction

The empirical description of topology and geometry of fault and fracture networks is an essential step in any investigation of the material properties of faulted and fractured media. The mechanical properties of fractured media such as stiffness or bulk rock strength differ significantly depending on the abundance of discontinuities. Rock stability is especially affected by the heterogeneity of fractures. Therefore, the analysis of the spatial arrangement of discontinuities allows for assessment of mechanical resilience and stability, for instance in underground mines (Wang, Jing, Su and Xie, 2017; Kong, Jiang, Shu, Sainoki and Wang, 2019). Furthermore, the characterization of fault and fracture networks represents a critical step in evaluating reservoir quality (Bauer, Krumbholz, Meier and Tanner, 2017) and assessing fluid flow properties (López and Smith, 1995; ?).

While faults and fractures are three dimensional discontinuities, we often rely on two-dimensional data as 3D datasets are difficult and too costly to obtain. Widely applicable characterizations of discontinuity patterns should therefore be based on the frequently available 2D trace maps. Depending on the area of interest (i.e. drill core, outcrop, or reservoir), the size of features to be analysed can range from millimetres up to several kilometres. Especially when considering the generic architecture of fault zones that comprises of a macroscopic core that is surrounded by an arrangement of mesoscopic features in the damage zone (Faulkner, Jackson, Lunn, Schlische, Shipton, Wibberley and Withjack, 2010; Sutherland, Toy, Townend, Cox, Eccles, Faulkner, Prior, Norris, Mariani, Boulton, Carpenter, Menzies, Little, Hasting, De Pascale, Langridge, Scott, Lindroos, Fleming and Kopf, 2012) the need for an analysis across scales becomes evident.

Key geometric properties of faults and fracture traces are their orientation, length, and sinuosity. The spatial arrangement can be characterized by determining density, intensity, and spacing, while the topological characterization assesses the connectivity and continuity (Sanderson and Nixon, 2015). The fractal geometry of the entire trace map, of the single lineaments as well as of the lineament centres and intersection points, can provide additional information on the spatial arrangement and the topology of the network (Roy, Perfect, Dunne and McKay, 2007).

Here we present an automated framework for 2D lineament analysis to characterize geometry, spatial arrangement, and topology. Geometric parameters are investigated in a statistical manner with maximum likelihood fitting to length distribution, and kernel density estimation of the principal orientations. In addition to the geometric properties, the spatial arrangement of the network is investigated, we apply window sampling to obtain densities, intensities, lineament centre to lineament centre spacing for the 2D datasets, and a classification of the spatial arrangement is obtained via box counting.

* This research was conducted as part of the CSIRO Deep Earth Imaging Future Science Platform

*Corresponding author

 uli.kelka@csiro.au (U. Kelka)

 <https://research.csiro.au/dei/people/ukelka/> (U. Kelka)

ORCID(s):

As density, intensity, and box counting dimension yield limited information on the topology we utilize a graph representation to quantify the connectivity of the network and characterize it in terms of the percolation threshold (Sanderson and Nixon, 2018). After converting the network into a graph, classical graph algorithms such as shortest path and maximum flow can be applied to further assess the fluid flow properties of the network. In addition to the analysis, the vector data forms the input for generating a finite element mesh on which detailed fluid flow simulations can be performed. The objectives of this contribution are: (i) introducing and testing an automatic data extraction tool for network analysis, (ii) demonstrate its performance on synthetic data, and (iii) showcase analysis of regional scale fault networks in South Australia.

2. Fault and Fracture Analysis

Fault and fracture analysis investigate the geometric properties, the spatial arrangement, and the topology of discontinuities in a network. The results can be utilized in hydrogeology (Partsinevelou, Lozios and Stournaras, 2011), reservoir characterization (Olson, Laubach and Lander, 2009; Abbasi, Sharifi and Kazemi, 2020), disposal systems for nuclear waste (Reeves, Parashar, Zhang and Rahman, 2012; Tsang, Neretnieks and Tsang, 2015), and can form the input for discrete fracture network models (Miyoshi, Elmo and Rogers, 2018). Often the acquired data is analysed to obtain statistical distributions and relationships between parameters (Zeeb, Gomez-Rivas, Bons and Blum, 2013, and references therein). Upscaling of parameters can be performed to assess overall physical properties of larger scales (e.g. equivalent permeabilities). Scaling relationships between certain parameters such as length and displacements in faults or length and aperture in fractures allow for obtaining additional information based on the statistical analysis of discontinuity sets. In the following we brief the key areas of the analysis: geometric properties, spatial arrangement, and topology.

2.1. Geometric Properties

Geometric properties refer to orientation, length distribution, and sinuosity of the components building up the network. The principal orientations of lineaments are related to stress states acting on the system whereas faults or fractures that form in a unique stress state are forming one group. Note that this is only valid for non-conjugate (orthogonal) features and the entire discontinuity network is often a result of all stress states that acted on the system (Maerten, Maerten, Lejri and Gillespie, 2016; Lei and Gao, 2018). The distribution of orientation in a network is therefore loosely related to the number, intensity, and sequence of deformation events.

The length distribution in natural discontinuity networks is found to follow either a log-normal, an exponential, or a power-law distribution (Zeeb et al., 2013; Healy, Rizzo, Cornwell, Farrell, Watkins, Timms, Gomez-Rivas and Smith, 2017). This is an essential parameter as the conductivity and storage are closely related to the length of a fracture, and the width of the damage zone scales with the length of a fault. These relationships are based on the correlation of length and displacement in faults (Gudmundsson, De Guidi and Scudero, 2013) and the scaling between aperture and length in fractures (Renshaw and Park, 1997; Mayrhofer, Schöpfer and Grasemann, 2019).

The sinuosity of a fault trace that can be observed on the surface is related to the stiffness (Young modulus) of the host rock and to the linkage between individual fault segments. Pronounced sinuosities of fault points towards heterogeneous stiffness along strike (Bridwell, 1975) and/or successive linkage of individual segments (Von Hagke, Kettermann, Bitsch, Bücken, Weismüller and Urai, 2019).

2.2. Spatial arrangement

The patterns of fault or fracture networks impact earthquake hazards, fluid flow properties, and resilience of rock masses, and their characterization requires several approaches (Laubach, Lamarche, Gauthier, Dunne and Sanderson, 2018b; Laubach, Hundley, Hooker and Marrett, 2018a). Analysing the spatial arrangement helps in quantifying the structural heterogeneity or anisotropy that governs permeability, resilience, and seismic properties of rock masses (cite). The major quantities of interest are density, intensity (Rohrbaugh Jr, Dunne and Mauldon, 2002; Zeeb et al., 2013; Watkins, Bond, Healy and Butler, 2015), spacing (?), and fractal dimension (Bour and Davy, 1999; Wilson, 2000, 2001).

Density and intensity are measures of the total frequency or cumulative length per unit area (Rohrbaugh Jr et al., 2002; Zeeb et al., 2013; Watkins et al., 2015) whereas high density values point towards structural complex zones which can represent areas comprising an enhanced vertical fluid flow component (Dimmen, Rotevatn, Peacock, Nixon and Nærland, 2017).

The spacing of discontinuities in an array is proposed to be a function of strength and stress state, and the mechanical thickness of the faulted layer (Soliva, Benedicto and Maerten, 2006; ?; ?). This measure helps to further quantify the heterogeneity of fault and fracture networks.

The abundance of faults and fractures across a wide range of scales may indicate that discontinuity pattern geometry is fractal (Bour and Davy, 1999, and references therein). One of the first attempts in describing the scale-invariance of natural fault systems was based on data obtained from the San Andreas fault and postulated a single parameter as governing the behaviour of the system, the fractal dimension $D - f$ (?). The fractal dimension can be considered as a metric for the characterization of the spatial arrangement of discontinuity networks (Walsh and Watterson, 1993; Idziak and Teper, 1996; Roy et al., 2007) but should be accompanied by other geometrical measures such as exponent of the length distribution (Bonnet, Bour, Odling, Davy, Main, Cowie and Berkowitz, 2001). A scaling relationship between the fractal dimension D_f and the exponent of the length distribution a was found to exist (Bour and Davy, 1999). The fractal geometry can help to quantitatively (describe?) size scaling and spatial clustering in fracture networks, which is crucial for predicting the hydrological and mechanical properties of fractured rock masses (Barton, 1995; Liu, Yu, Jiang, Wang and Li, 2017).

2.3. Topology

In contrast to the purely geometrical parameters such as length, orientation and spacing, which can be expressed in physical units, topology is an abstract way of analysing the relationship and arrangement of objects in space. In networks, the measures of topology classify connectivity and sparsity of models (Kincaid and Phillips, 2011).

The topological network analysis of natural discontinuity networks yields information on the partition of the solid into separated blocks (Husby et al., 1997) and on the connectivity (Andresen, Hansen, Le Goc, Davy and Hope, 2013). Topology represents a dimensionless measure of connectivity and arrangement of natural discontinuity networks where the objects are represented as the edges and vertices of a graph (Morley and Nixon, 2016; Sanderson, Peacock, Nixon and Rotevatn, 2019). In recent years the applicability of topological network analysis to fault and fracture networks has gained interest (Sanderson et al., 2019; ?) as the information on connectivity is directly related to fluid flow properties and permeability of natural networks (Manzocchi, 2002; Renard and Allard, 2013).

In addition to assessing flow properties, the topology of discontinuity networks also represents a metric that helps characterizing types of fault and fracture arrays (Morley and Nixon, 2016). As the number of intersections in a network that results from a sequence of different deformation events is generally larger than in single-phase deformation, topographical concepts also yield information on the development of an arrangement of faults (?).

2.4. Existing software

We would be remiss should we not mention the existing commercial and open-source software packages for fault and fracture analysis. A well-established commercial tool is `textscfracman` which allows for analysis and modelling of fracture networks (Fra, 2011). A variety of open-source frameworks for fracture analysis exist and we name a few of the latest: for acquisition and processing of digital outcrop data, `Digifract` was built on top of QGIS (Hardebol and Bertotti, 2013). `FraNep` is an ExcelTM based tool written in Microsoft Visual BasicTM for statistical evaluation of 2D fracture networks (Zeeb et al., 2013). A Matlab toolbox for fracture pattern quantification is `FracPAQ` (Healy et al., 2017). One of the latest contributions to open-source fault and fracture analysis is the ArcGIS toolbox `NetworkGT` (Nyberg, Nixon and Sanderson, 2018). Our approach to topological characterization is mainly based on the work of (Sanderson and Nixon, 2018; Sanderson et al., 2019) and therefore comparable to (Nyberg et al., 2018).

3. FracG workflow

Our framework utilizes the `gdal` library (Contributors, 2018) for handling geospatial data, boost libraries for the geometrical (Gehrels, Lalande, Loskot and Wulkiewicz, 2016) and topological analysis (Siek, Lee and Lumsdaine, 2001), the GNU Scientific Library (GSL) is utilized for statistic and mathematical analysis (Gough, 2009), and automatic 2D finite mesh generation is realized by using `gmsh` API (Gehrels et al., 2016). As the statistical analysis of large discontinuity networks that are for instance obtained by unmanned aerial vehicles (UAV) can be computationally expensive. We therefore choose C++ as the programming environment and parallelized the most demanding part of the analysis with openMP (Dagum and Menon, 1998).

The network analysis is based on reading in an ESRI Shapefile containing lineament locations in a projected coordinate system. Note that geographic coordinate systems are currently not supported and the user will have to reproject

his data into an projected reference system prior to the analysis. A gneric workflow of a complet anlaysis also including rater data is outlined in figure 2.

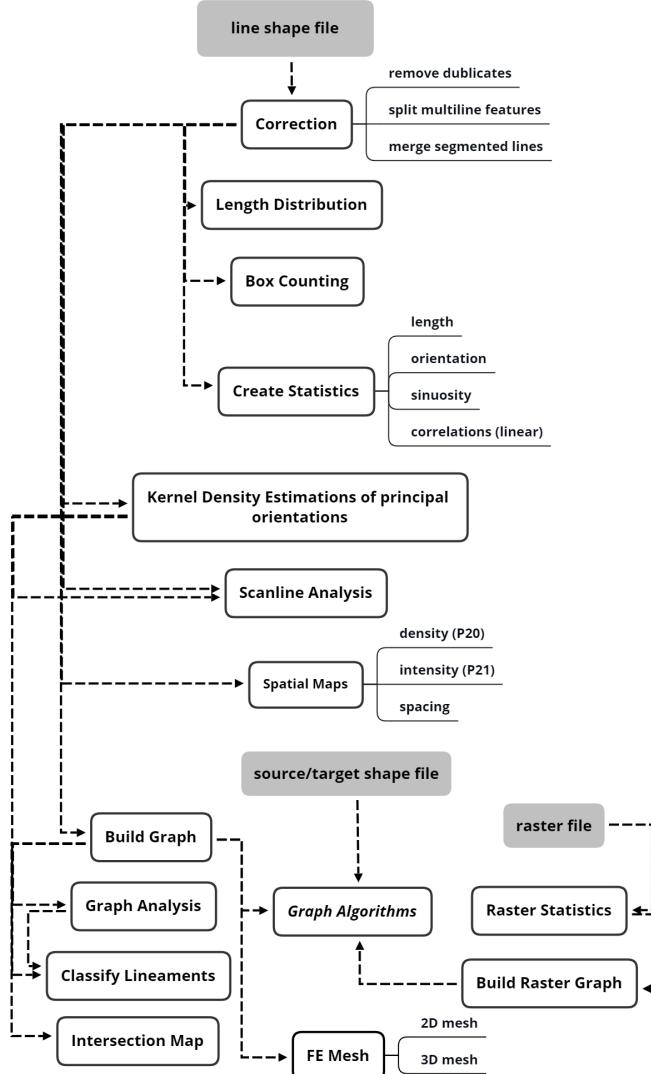


Figure 1: Schematic work flow of *FracG* showing the different possible analysis. Grey shaded boxes represent data that needs to be passed to the program, the other boxes represent the different functions with associated outputs or performed tasks. The dotted arrows indicate how the analysis progresses from initial input to the different products and how the respective functions are interlinked.

3.1. Data import and correction

Data on geological faults often represents a compilation of different datasets where the compiled data frequently exhibits flaws that would prevent reliable analysis of geometric properties. Therefore, the first step after importing the data format is to correct these flaws. After checking for a projected coordinate system, the data is read in and multi-lines are split into individual features. The subsequent step is the correction of the data is to merge falsely segmented lineaments. The features are merged based on a user-defined threshold representing the maximum distance between line-tips, and a maximum difference in orientation between adjacent line features is set to a user-defined range . This step often significantly reduces the number of separate digitized discontinuities and is crucial for performing reliable statistical analysis. After the correction the lineament set is also checked for potential duplicates and the corrected data can be written into a shape file.

4. Geometrical Analysis

We investigate the distributions of length, orientation, and sinuosity. At the current stage, the analysis is performed in Cartesian coordinates allowing for quick calculations but limiting the spatial accuracy for large scale data, for instance on a continental scale.

4.1. Length distribution

We obtain a statistical model for the length distribution and determine whether the data follows an exponential, log-normal, or a power-law distribution. We adapted the methods of (Clauset, Shalizi and Newman, 2009) for this problem the reader is referred to that work for further details. In the following we brief the general approach to obtaining the best fit of the empirical to a specific model:

First, we derive the model parameters for each distribution by maximum-likelihood methods and calculate the Kolmogorov-Smirnov (*KS*) statistic between the data and each model. For each model, it is possible that there is a minimum value in the empirical distribution such that it is only accurately described by the statistical model above that minimum value. Such a minimum cut-off is chosen by determining the model parameters and evaluating the *KS* statistic for each data point in the distribution and selecting the point with the minimum *KS* statistic.

The suitability of the model for describing the empirical data is determined by comparing the *KS* statistic of the model and data, against the *KS* statistics calculated comparing sample distributions randomly drawn from the model and the new model which is generated from that random sample. In order to capture the characteristics of the original distribution, sample data points are drawn from either the candidate model or the below-minimum values of the empirical dataset, with a probability equal to the portion of the data points in the dataset that are above the minimum cut-off.

The number of samples, and therefore executions of the following steps, is derived from the maximum error (*err*) we allow for out *p*-value metric as:

$$N = \frac{1}{4}err^2 \quad (1)$$

By this procedure, synthetic datasets of the same size as the original data are generated and the model parameters are derived from this synthetic dataset via maximum-likelihood estimates. We count the number of random sample distributions that result in a *KS* value that is larger than the *KS* value of the original dataset as *c*. The *p*-value that determines whether a model will be rejected is derive from this counter as:

$$p = \frac{c}{N} \quad (2)$$

Higher *p*-values suggest that the model fits the data better. If there is more than one acceptable model, we proceed to determining which distribution fits the empirical data best. The goodness-of-fit is tested via pairwise log-likelihood comparisons of the candidate models. The minimum datapoint used in this comparison is the largest minimum value of the two distributions being compared. The log-likelihood ratio (*R*) is calculated to determine which model is the better fit (table 1). We determine the reliability of this comparison using a second *p*-value type, calculated here as:

$$p = erfc \left(\frac{|R|}{\sqrt{2N}} \right) \quad (3)$$

Here a lower value suggests more reliable comparison, and we consider the threshold for a meaningful comparison to be 0.1. After this procedure we report the type and parameters of the distribution(s) that best fit the empirical data.

In figure 3 we demonstrate the automatic fitting based on randomly created data synthetic data that follows an exponential (a), a log-normal (b) and a power law distribution(c).

4.2. Box-Counting dimension

The “Minkowski-Boulingard” dimension is determined via box counting. For a detailed description of the box counting method in fracture networks we refer to (?). We will briefly outline our implementation of the box-counting

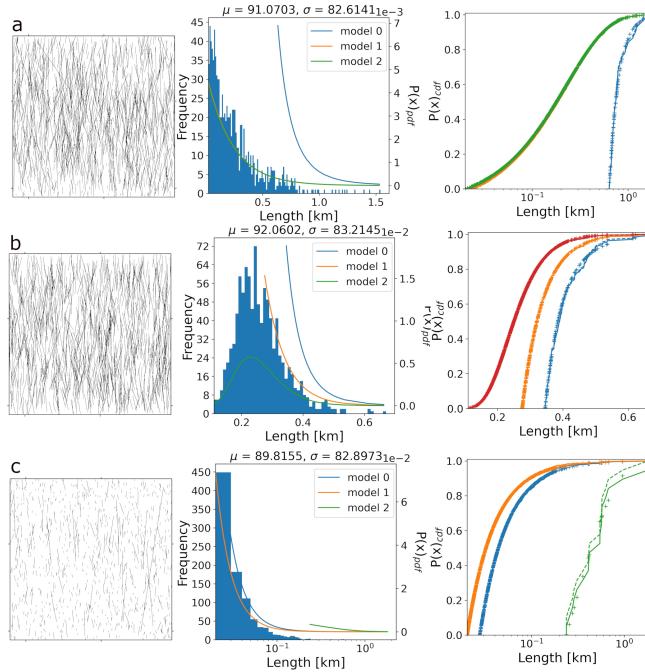


Figure 2: Example of fitting exponential, log-normal, or power law models to synthetic data comprising 1000 lineaments whose length follow one of the three distributions. In each row of this figure the synthetic data (right), the probability density functions of the models compared to the histogram (centre), and the cumulative distribution functions of the models are shown.

a synthetic exponential: Three models are fitted for different data ranges. Model 0 follows a power law distribution ($c: 0.0067; \alpha: 5.373$ above 634.139 m representing 74 lineaments). Model 1 follows a exponential distribution with $\lambda: 0.00439$ for lineament longer than 23.1241m representing 994 lines. Model 3 fits an exponential distribution to the complete data range with $\lambda: 0.004367$.

b synthetic log-normal: Three models are fitted to different ranges of this randomly created dataset. Model 0 follows a Power law distribution with $c: 0.0188$ and $\alpha: 7.5243$ for lineaments longer than 345,705 corresponding to 134 of the 1000 lines. Model 1 fits an exponential model with $\lambda: 0.0153$ to all lines longer than 275.256 which are 374 lines. Model 2 represents a log-normal model with $\mu: 5.521$ and $\sigma: 0.29$ for the entire range of lines in the dataset.

c synthetic power-law: Three models are fitted to this synthetic dataset. Model 0 represents a power-law fit with $c: 0.0544$ and $\alpha: 2.493$ for lines longer than 27.4332 which are 639 of the dataset. Model 1 fits a power-law distribution to the entire range of the model with parameters $c: 0.0738$ and $\alpha: 2.477$. Model 3 represents an exponential fit to lines above 237.597 which are 19 of the total 1000.

algorithm, by utilising quad trees. The maximum scale is represented by boxes with a side length equal to the longest lineament in the set whereas the smallest scale is derived from an arbitrary (user-defined?) number of recursions. During each recursion the area covered by each box is quartered. The maximum scale is represented by boxes with a side length equal to the longest lineament in the set whereas the smallest scale is derived from an number of recursions. During each recursion the area covered by each box is quartered.

The first layer of boxes is created, covering the entire bounding box. Note that the origin of these covering boxes may be iteratively offset to evaluate the uncertainty of the fractal dimension. In the next step, the number of non-empty boxes (n) is determined. At each scale, where a lineament overlaps a box, that box is marked as filled, and we compare the lineament against that box's four quadrants. When a lineament overlaps with a quadrant, the comparison is performed recursively using the quadrant and the portion of the lineament that overlaps with that quadrant. This recursive refinement is performed until the minimum scale is reached. In this manner, the count of non-empty boxes is obtained for different sizes. For each scale, we record the side length of the boxes (s) and the number of boxes that are filled (n). We use least-squares fitting to fit a line of $\log(n)$ as a function of $\log(s)$ and report its slope as the fractal dimension. (TODO: double-check what we do with the box counts after they are reported.)

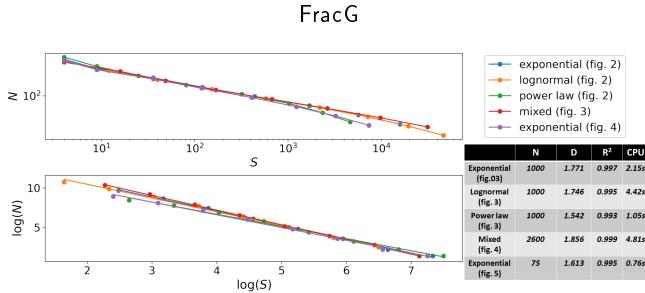


Figure 3: Results of the box counting for some of the synthetic networks shown in this manuscript. As the maximum and minimum box size is related to the largest and smallest feature in the data set, the computational time is not only a function of the total number (density) of entities but also strongly depends on the range of the feature length. Due to the implementation of the quad-tree method computational time is highest for data that fills the area of interest.

4.3. Principal Orientations

The orientation of a lineament is calculated from the lineament tips. The angle of orientation is expressed as a value between 0deg and 180deg. This data can be used to distinguish different lineament sets and identify major structural trends in an area. An estimated probability density function (*epdf*) of the angle distribution is produced using Kernel Density Estimation (*KDE*, **table 1**), with a Gaussian kernel. We attempt to fit a multiple Gaussian to the *epdf* by first attempting to fit between one and ten gaussians. For each number of Gaussians, we calculate the Akaike Information Criterion (*AIC*), and choose the number of Gaussians that results on the lowest *AIC*. The non-linear function fitting algorithm from the GSL is used to fit each number of Gaussian to the *epdf*. The initial parameters for the fitting are determined from the inflection points of the *epdf*. We create pairs of falling, then the next rising, zero-crossings its second derivative. For each pair, the position is set as the mid-point of the two crossings, the amplitude is the value of the *epdf* at that position, and the width is set as the distance between the zero-crossings. If we find more pairs of zero-crossings than the number of Gaussians we are currently trying to fit, the *epdf* is increasingly smoothed until it has the desired number of crossings. This smoothing is reset for each number of Gaussians. If the number of crossings we find is equals to the number of Gaussians we are currently evaluating, then we calculate the residual between the *epdf* data and the pdf modelled by the fitted Gaussians and use this residual as the pdf data when trying to fit greater numbers of Gaussians. If we are still unable to find additional pairs of inflection points after calculating the residuals, then we stop trying to fit additional Gaussians and only consider the values that we have already fit.

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K \frac{x - x_i}{h} \quad K = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} \quad \int h_h(x) dx = 1 \quad (4)$$

4.4. Sinuosity

The sinuosity is defined as the ratio between the length of a feature and the geometrical distance between its end points.

Table 1
Geometric parameters derived during the analysis. .

Parameter	Expression
Sinuosity (<i>s</i>)	$s = \frac{D}{L}$
Orientation (α)	$\alpha = \text{atan} \left(\frac{x_1 - x_2}{y_1 - y_2} \right) \frac{180}{\pi}$
Orientations (<i>Nb</i>)	$Nb = \forall x : f'_h(x) = 0 \wedge f''_h(x) < 0$
Length distribution (d)	$R = \text{sumln} (pd f_a(x)) - l (pd f_b(x))$
	$\sigma = \sqrt{\frac{1}{N} \sum \left((\ln(pd f_a(x)) - (pd f_b(x))) - \frac{R}{N} \right)^2}$
Density	$P_{10} = \frac{N}{L}; P_{20} = \frac{N}{A}; P_{20_c} = \frac{N}{2\pi r^2}$
Intensity	$S = \frac{1}{P_{10}}; P_{21} = \frac{\sum l}{A}; P_{21_c} = \frac{\pi r^2 n}{2 m}$

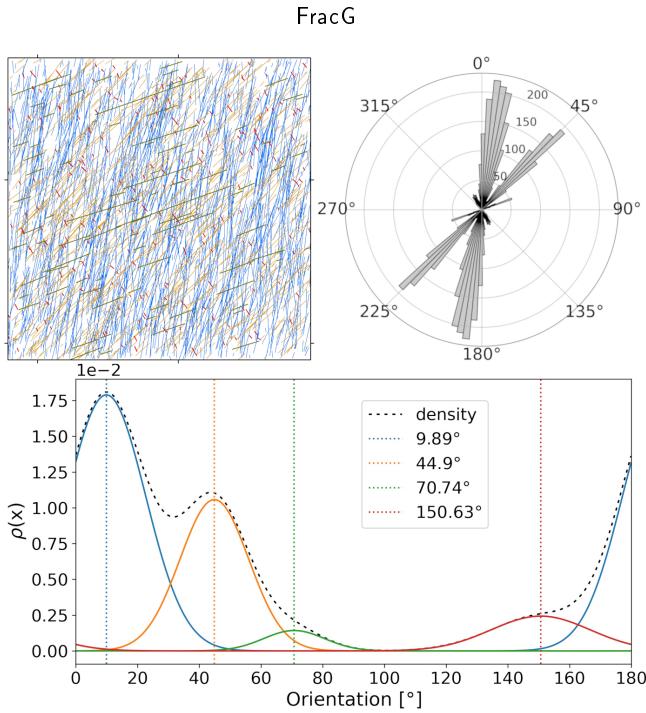


Figure 4: Example of fitting Gaussian models to a synthetic dataset comprising 2600 features that are generated for four different directions with variable numbers and variances. The principal directions visible by in the rose diagram can automatically be obtained via iteratively fitting models comprising different numbers of Gaussian, the amplitude of the Gaussian correspondent to the number of lines that fit into this distribution. Note that in lower plot of this figure the areas below the Gaussian is chosen to add up to one and the area of each Gaussian represent the frequency of the elements belonging to the particular Gaussian.

5. Spatial arrangements

With spatial arrangement we refer the density or length per unit area, the Minkowski-Boulingard dimension, and the topology.

5.1. Density, Intensity, and Spacing

The characterization of the networks in terms of density, intensity, and spacing can be performed either with scan lines or through window sampling (Zeeb et al., 2013; Watkins et al., 2015; Sanderson et al., 2019). With these sampling methods the relative abundance of fractures can be expresses as density (P), spacing (S) or intensity (I). Based on the sampling method, linear intensity (P_{10}) is obtained though line scan sampling and lineament intensity (P_{20}) and length intensity (P_{21}) are obtained from window sampling.

5.1.1. Scanline sampling

The metrics obtained by scanline sampling are the average spacing (\bar{S}) and the liner intensity (P_{10}). In order to minimize the sampling bias, numerous scanlines need to be analysed that are oriented perpendicular to the principal orientations of the analysed features. Therefore, scanline sampling in FracG can only performed if the principal orientation were obtained (see section 4.3). The start points of the sampling lines are randomly chosen within the convex hull of the domain. This ensures that the analysis is focused on the areas comprising features while omitting void areas. The orientation of the scanline by randomly choosing a orientation that is perpendicular to one of the principal orientations. This minimizes orientation dependent bias of scanline sampling to some extend but cannot address bias caused by truncation or censoring effects that would already be present in the initial data. Intensity (P_{10}) and average spacing (\bar{S}) are calculated as:

$$P_{10} = \frac{I}{L} \quad \bar{S} = \frac{l}{I - 1} \quad (5)$$

Where I is the total number of intersection, L is length in m, and l is the distance between consecutive points along the scaline. For every scanline the underlying principal orientation, the start- and endpoint coordinates, length, number of intersections, intersection intensity (P_{10}) and the average spacing are reported in a table. The user can define the total number of scalines and the desired spacing between lines of the same orientation. For large sample-line numbers FracG will attempt to fill the domain with lines that are close to the horizontal extend of the rectangular area while maintaining the predefined minimum spacing. After each failed attempt of constructing a line either the minimum length of the scanline or the minimum spacing is reduced. If after $1e^5$ failed attempt to construct a line, the program will terminate.

In figure 5 we show 50 scanlines generated for a synthetic dataset comprising two principal orientation following an exponential length distribution.

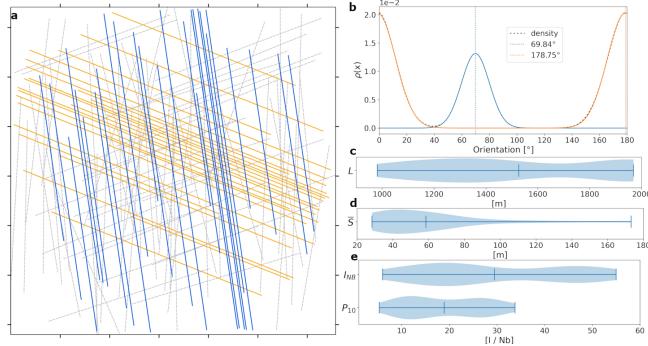


Figure 5: a 50 scanlines that were randomly created following a uniform random distribution. The two sets of sample lines are constructed perpendicular to one of the principal orientation that were obtained via fitting Gaussians to the KDE. c shows the length distribution of scanlines, d shows the distribution of average spacing, and e shows the intersection numbers and linear intensities of the 50 scanlines.

5.1.2. Window sampling

Rectangular Window sampling is performed for obtaining density (total number, P_{20}), intensity (cumulative length, P_{21}), and distances per unit area for the domain. Two types of distance maps are created: the minimum distance of every pixel centre to the nearest line, and the distance to the nearest line centre. Every sampling window is a pixel in the created raster file whereas the window or pixel size can be defined by the user and is initially set to 1000m by 1000m. For a better visualization reampling to a 10^th of

In figure 5 we show the four maps types obtained for a synthetic networks comprising 75 elements (same networks as in figure 4. The left image column show the data created by window sampling with a pixels size of .

The right a map is re-sampled to a pixel size of one 10^{th} of the window size using cubic B-spline approximation with a 4 by 4 kernel interpolation. Note that the pixel values are projected back to the initial range after re-sampling to avoid unnatural odd number for instance for the density maps.

6. Topological Analysis

We follow the approach described by Sanderson and Nixon (2018); Sanderson et al. (2019) to obtain the topology metrics from a georeferenced graph. This data structure can be linked to raster data and utilized for assessing additional network properties by applying graph algorithms. In a 2D projection, the system can be presented as a set of vertices and edges. This graph yields information on geometry as well as on topology and allows for characterization of the network in terms connectivity and percolation (Manzocchi, 2002; ; Sanderson and Nixon, 2015, 2018). For the graph representation the vertices are labelled depending on the number of edges incident with this vertex, called the degree of the vertex. Three different degrees of vertices exist: isolated tips (1), junctions (3), and crossings (4). Even if the fault traces are not connected at the surface, a hydraulic connectivity between neighbouring faults might exist through the fault damage zone (figure 2 a). Damage zone geometries differ depending on the location relative to (distance from...?) the fault core (Peacock, Dimmen, Rotevatn and Sanderson, 2017). Along a single fault the overall fracture density in the wall damage (zone?) (Wdz) and tip damage zone (Tdz) will have different geometries and fracture densities. In

a fracture network, damage zones at junctions (Ydz) or crossings (Xdz) will comprise higher fracture densities. Via linking damage zones (Ldz) a hydraulic connectivity between neighbouring faults can be established even if the 2D fault traces do not intercept. The lineaments are converted into an undirected adjacency list with bundled properties (maybe rephrase, in order to focus on what is important to the reader). Geographic coordinates of the edges (line strings) and the vertices (points) are directly assigned to graph. The next step is to locate the intersection points of the lineaments, split the traces at these intersection points, and add vertices at these intersections. After this splitting but before the graph's analysis, the isolated tips whose lengths are below a defined threshold will be removed from the graph. This ensures that y-intersections are not falsely interpreted as x-intersections. Our framework accounts for this by defining a minimum distance threshold, where lineaments that are within this distance of each other are considered to intersect even if they do not overlap.

6.1. Graph representation

At the beginning of the analysis the consistency of the created graph is tested via Boyer-Myrvold Planarity Test. This test ensures that the network geometry is captured correctly, and Euler's formula applies. Subsequently, the graph and connected components comprising at least a certain number of edges that can be defined by the user (this sentence is incomplete). The percolation state can be evaluated in a tertiary diagram (MANZOCCHI, 2002) by utilizing the reported values. The weights assigned to the edges are the geographical length allowing for further spatial investigation of the network. The area that is occupied by the network is determined by the convex hull around the graph which minimized the bias for area-dependent calculations, such as line and branch frequency. The parameters derived from the graph analysis are reported as a tab-separated text file and are summarized in table 2.

Table 2

Parameters obtained from the graph. Number of line tips (N_l), number of Y-nodes (N_y), number of X-Nodes (N_x), branch types: IX, IY, II, YY, XX, number of lines (NL), length of branch(l_b), area of convex hull around graph (A), node degree (deg(v)).

Parameter	Expression
Number of Vertices (N _v)	N _v = N _i + N _y + N _x
Number of Branches (N _b)	N _B = IX + IY + YX + YY + XX + II
Connecting Nodes (N _c)	N _c = NX + NY
Average Branch Length (l̄ _B)	l̄ _B = $\frac{\sum_{i=0}^n l_{B_i}}{N_B}$
Connecting Node Frequency (f _{NC})	f _{NC} = $\frac{\sum N_y + \sum N_x}{N_v}$
Dimensionless Intensity (B ₂₂)	B ₂₂ + P ₂₁ L _B
Average degree (deg _{av})	deg _{av} + $\frac{\sum deg(v)}{N_v}$
Average connection per line (C _{av})	C _{av} = $\frac{2N_x + N_y}{N_v}$
Number of faces (f)	f = 2N _v + N _B

6.1.1. Graph Algorithms

graph analysis that are allied for obtaining more topological information about the networks are minimum spanning tree, shortest path, and betweenness centrality. We consider solving maximum flow as a separate problem and describe this in greater detail in the subsequent section.

A spanning tree is a tree structure that maintains the overall connectivity of the vertices. A single undirected graph can have several of such structures whereas the minimum spanning tree is one with the lowest sum of edge weights.

The shortest path algorithm is applied to a weighted undirected graph where the edge weights can either be the length of the edge, the centre value, the mean value of the edge, or the profile along the edge. The algorithm we utilize is the Dijkstra Shortest Path from the boost graph library.

6.2. Maximum Flow

We utilize the “Bojkov-Kolmogorov Maximum Flow algorithm” of the boost graph library which requires the creation of a directed graph with two directed edges for each undirected edge. The value associated with each directed edge is the maximum capacity of fluid flow between its start and end vertices, in each direction of its corresponding

undirected edge (this could be phrased better). We implemented three different functions to derive edge capacities based on their length, orientation, or a combination of length and orientation. The direction of the flow is derived from vertex weights of an edge as the flow is always from high to lower values, taking also into account that the flow cannot reach higher than the source vertex.

The vertex weights (we're not really using these as "weights", maybe we should call them the height of each vertex) can either be derived from a raster file (e.g. DEM) or from horizontal or vertical gradients (I need to check what we're doing here – I've only seen using raster file as a DEM, and in any case those gradients will also come from a raster). The maximum and minimum values of the gradients must be defined by the user and a liner (is this the correct word?) gradient between the two values is generated in the envelope around the graph (is the envelope around the graph or the edge? Also, if it is around the graph, call it the "region of interest" to be consistent with what is said above about using a raster file). The capacities are derived as described in table 3.

Table 3

Parameters obtained from the graph. Number of line tips (NI), number of Y-nodes (NY), number of X-Nodes (NX), branch types: IX, IY, II, YY, XX, number of lines (NL), length of branch(lb), area of convex hull around graph (A), node degree (deg(v)).

Capacity dependency	Expression
length (cap_l)	$cap_l = \frac{\kappa_f \sqrt{l}}{\frac{1}{2} \kappa_m l}$
orientation (cap_o)	$cap_o = \frac{1}{\sigma \sqrt{2\pi} \exp^{-0.5 \frac{a-\mu}{\sigma^2}}} \times 1000$
length and orientation (cap_{lo})	$cap_{lo} = cap_l \cdot cap_o$

7. Raster data

The analysis can be extended by adding raster data sets to the vector data. The rectangular area of interest is defined by the extent of each raster file and noData values are set to 0. The general statistics of the raster data such as minima, maxima, mean, and standard deviation are reported in a csv file together with the pixel values along strike and perpendicular to strike at the centre of the lines. The length of the cross line is 2.5 pixel resolution and representing the sampling (d_s) distance utilized for most of the raster analysis.

Lineaments which cross the borders of the raster will be cut to the raster area and marked as edge-nodes (E-nodes). Lineaments containing an edge node will be excluded from the statistical analysis of lineament length.

After generating a graph clipped to the extent of the raster file, the statistical topological analysis previously described will be performed only for the area contained in the raster. Note that the extent of the raster files that are added to the analysis do not have to be consistent, but the geographical projection must be the same for the vector and any raster data that is added.

The values of the raster file are stored as edge and vertex weights. Data is collected at every vertex, at the centre (centre of what?), and along each edge. In addition, the gradient across the centre of the edges, the mean gradient across the segments (profile) and the gradient along the branches (before I forget, do we formally state what our definition of a branch is?) is calculated. Each value is stored as an edge weight which allows for performing additional analysis on the graph (Need to double check this - each edge in a graph only has a single weight. If we use different weights in different places, then we should say that we store these properties, and mention below when each property is used as a weight). For instance, if the raster data is a digital elevation model (DEM), the path with the maximum topographic gradient can be calculated based on the maximum flow algorithm.

FIGURE 10

8. Automatic Meshing: 2D with local refinement

A finite element mesh can be created from the vector data utilizing the open-source mesh generator gmsh. to generate 2D meshes conforming to the discontinuity traces that are represented by side sets. Discretizing the features as side sets allows for applying a mixed dimensional approach where faults or fractures are represented by elements of a lower dimensional order compared to the domain they are hosted in (Poulet, Lesueur and Kelka, 2021). Mesh-refinement is a function of distance to the sidesets. FIGURE 11

9. Example: Central Gawler Craton, South Australia

FIGURE 12

10. Conclusion

11. Discussion

A. My Appendix

Appendix sections are coded under \appendix.

\printcredits command is used after appendix sections to list author credit taxonomy contribution roles tagged using \credit in frontmatter.

CRediT authorship contribution statement

Ulrich Kelka: Conceptualization of this study, Methodology, Software, Data curation, Writing - Original draft preparation. **Stefan Westerlund:** Methodology, Software, Data curation, Writing - Original draft preparation. **Luk Peeters:** Data curation, Writing - Original draft preparation.

References

- , 2011. Fracman7 user documentation.
- Abbasi, M., Sharifi, M., Kazemi, A., 2020. Fluid flow in fractured reservoirs: Estimation of fracture intensity distribution, capillary diffusion coefficient and shape factor from saturation data. *Journal of Hydrology* 582, 124461.
- Andresen, C.A., Hansen, A., Le Goc, R., Davy, P., Hope, S.M., 2013. Topology of fracture networks. *Frontiers in Physics* 1, 7.
- Barton, C.C., 1995. Fractal analysis of scaling and spatial clustering of fractures, in: *Fractals in the earth sciences*. Springer, pp. 141–178.
- Bauer, J.F., Krumbholz, M., Meier, S., Tanner, D.C., 2017. Predictability of properties of a fractured geothermal reservoir: the opportunities and limitations of an outcrop analogue study. *Geothermal Energy* 5, 1–27.
- Bonnet, E., Bour, O., Odling, N.E., Davy, P., Main, I., Cowie, P., Berkowitz, B., 2001. Scaling of fracture systems in geological media. *Reviews of geophysics* 39, 347–383.
- Bour, O., Davy, P., 1999. Clustering and size distributions of fault patterns: Theory and measurements. *Geophysical Research Letters* 26, 2001–2004.
- Bridwell, R., 1975. Sinuosity of strike-slip fault traces. *Geology* 3, 630–632.
- Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. *SIAM review* 51, 661–703.
- Contributors, G., 2018. Gdal/ogr geospatial data abstraction software library. Open Source Geospatial Foundation .
- Dagum, L., Menon, R., 1998. Openmp: an industry standard api for shared-memory programming. *Computational Science & Engineering*, IEEE 5, 46–55.
- Dimmen, V., Rotevatn, A., Peacock, D.C., Nixon, C.W., Nærland, K., 2017. Quantifying structural controls on fluid flow: Insights from carbonate-hosted fault damage zones on the maltese islands. *Journal of Structural Geology* 101, 43–57.
- Faulkner, D., Jackson, C., Lunn, R., Schlische, R., Shipton, Z., Wibberley, C., Withjack, M., 2010. A review of recent developments concerning the structure, mechanics and fluid flow properties of fault zones. *Journal of Structural Geology* 32, 1557–1575.
- Gehrels, B., Lalande, B., Loskot, M., Wulkiewicz, A., 2016. Boost geometry library.
- Gough, B., 2009. GNU scientific library reference manual. Network Theory Ltd.
- Gudmundsson, A., De Guidi, G., Scudero, S., 2013. Length-displacement scaling and fault growth. *Tectonophysics* 608, 1298–1309.
- Hardebol, N., Bertotti, G., 2013. Digifract: A software and data model implementation for flexible acquisition and processing of fracture data from outcrops. *Computers & Geosciences* 54, 326–336.
- Healy, D., Rizzo, R.E., Cornwell, D.G., Farrell, N.J., Watkins, H., Timms, N.E., Gomez-Rivas, E., Smith, M., 2017. Fracpaq: A matlab™ toolbox for the quantification of fracture patterns. *Journal of Structural Geology* 95, 1 – 16. URL: <http://www.sciencedirect.com/science/article/pii/S0191814116302073>, doi:<https://doi.org/10.1016/j.jsg.2016.12.003>.
- Idziak, A., Teper, L., 1996. Fractal dimension of faults network in the upper silesian coal basin (poland): Preliminary studies. *pure and applied geophysics* 147, 239–247.
- Kincaid, R.K., Phillips, D.J., 2011. Network topology measures. *Wiley Interdisciplinary Reviews: Computational Statistics* 3, 557–565.
- Kong, P., Jiang, L., Shu, J., Sainoki, A., Wang, Q., 2019. Effect of fracture heterogeneity on rock mass stability in a highly heterogeneous underground roadway. *Rock Mechanics and Rock Engineering* 52, 4547–4564.
- Laubach, S., Hundley, T., Hooker, J., Marrett, R., 2018a. Spatial arrangement and size distribution of normal faults, buckskin detachment upper plate, western arizona. *Journal of Structural Geology* 108, 230–242.
- Laubach, S.E., Lamarche, J., Gauthier, B.D., Dunne, W.M., Sanderson, D.J., 2018b. Spatial arrangement of faults and opening-mode fractures. *Journal of Structural Geology* 108, 2–15.
- Lei, Q., Gao, K., 2018. Correlation between fracture network properties and stress variability in geological media. *Geophysical Research Letters* 45, 3994–4006.

- Liu, R., Yu, L., Jiang, Y., Wang, Y., Li, B., 2017. Recent developments on relationships between the equivalent permeability and fractal dimension of two-dimensional rock fracture networks. *Journal of Natural Gas Science and Engineering* 45, 771–785.
- López, D.L., Smith, L., 1995. Fluid flow in fault zones: analysis of the interplay of convective circulation and topographically driven groundwater flow. *Water resources research* 31, 1489–1503.
- Maerten, L., Maerten, F., Lejri, M., Gillespie, P., 2016. Geomechanical paleostress inversion using fracture data. *Journal of structural Geology* 89, 197–213.
- Manzocchi, T., 2002. The connectivity of two-dimensional networks of spatially correlated fractures. *Water Resources Research* 38, 1–1.
- Mayrhofer, F., Schöpfer, M.P., Grasemann, B., 2019. Universal and nonuniversal aperture-to-length scaling of opening mode fractures developing in a particle-based lattice solid model. *Journal of Geophysical Research: Solid Earth* 124, 3197–3218.
- Miyoshi, T., Elmo, D., Rogers, S., 2018. Influence of data analysis when exploiting dfn model representation in the application of rock mass classification systems. *Journal of Rock Mechanics and Geotechnical Engineering* 10, 1046–1062.
- Morley, C., Nixon, C., 2016. Topological characteristics of simple and complex normal fault networks. *Journal of Structural Geology* 84, 68–84.
- Nyberg, B., Nixon, C.W., Sanderson, D.J., 2018. Networkgt: A gis tool for geometric and topological analysis of two-dimensional fracture networks. *Geosphere* 14, 1618–1634.
- Olson, J.E., Laubach, S.E., Lander, R.H., 2009. Natural fracture characterization in tight gas sandstones: Integrating mechanics and diagenesis. *AAPG bulletin* 93, 1535–1549.
- Partsinevelou, A., Lozios, S., Stournaras, G., 2011. Fracture pattern description and analysis of the hard rock hydrogeological environment in naxos island, hellas, in: *Advances in the Research of Aquatic Environment*. Springer, pp. 45–52.
- Peacock, D., Dimmen, V., Rotevatn, A., Sanderson, D., 2017. A broader classification of damage zones. *Journal of Structural Geology* 102, 179–192.
- Poulet, T., Lesueur, M., Kelka, U., 2021. Dynamic modelling of overprinted low-permeability fault cores and surrounding damage zones as lower dimensional interfaces for multiphysics simulations. *Computers & Geosciences* 150, 104719.
- Reeves, D.M., Parashar, R., Zhang, Y., Rahman, R., 2012. Hydrogeologic characterization of fractured rock masses intended for disposal of radioactive waste. *INTECH Open Access Publisher*.
- Renard, P., Allard, D., 2013. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources* 51, 168–196.
- Renshaw, C., Park, J., 1997. Effect of mechanical interactions on the scaling of fracture length and aperture. *Nature* 386, 482–484.
- Rohrbaugh Jr, M., Dunne, W., Mauldon, M., 2002. Estimating fracture trace intensity, density, and mean length using circular scan lines and windows. *AAPG bulletin* 86, 2089–2104.
- Roy, A., Perfect, E., Dunne, W.M., McKay, L.D., 2007. Fractal characterization of fracture networks: An improved box-counting technique. *Journal of Geophysical Research: Solid Earth* 112.
- Sanderson, D.J., Nixon, C.W., 2015. The use of topology in fracture network characterization. *Journal of Structural Geology* 72, 55–66.
- Sanderson, D.J., Nixon, C.W., 2018. Topology, connectivity and percolation in fracture networks. *Journal of Structural Geology* 115, 167–177.
- Sanderson, D.J., Peacock, D.C., Nixon, C.W., Rotevatn, A., 2019. Graph theory and the analysis of fracture networks. *Journal of Structural Geology* 125, 155–165.
- Siek, J.G., Lee, L.Q., Lumsdaine, A., 2001. The boost graph library: User guide and reference manual, portable documents. Pearson Education.
- Soliva, R., Benedicto, A., Maerten, L., 2006. Spacing and linkage of confined normal faults: importance of mechanical thickness. *Journal of Geophysical Research: Solid Earth* 111.
- Sutherland, R., Toy, V., Townend, J., Cox, S., Eccles, J., Faulkner, D., Prior, D., Norris, R., Mariani, E., Boulton, C., Carpenter, B., Menzies, C., Little, T., Hasting, M., De Pascale, G., Langridge, R., Scott, H., Lindroos, Z.R., Fleming, B., Kopf, A., 2012. Drilling reveals fluid control on architecture and rupture of the Alpine fault, New Zealand. *Geology* 40, 1143–1146. URL: <https://doi.org/10.1130/G33614.1>, doi:10.1130/G33614.1, arXiv:<https://pubs.geoscienceworld.org/geology/article-pdf/40/12/1143/3543987/1143.pdf>.
- Tsang, C.F., Neretnieks, I., Tsang, Y., 2015. Hydrologic issues associated with nuclear waste repositories. *Water Resources Research* 51, 6923–6972.
- Von Hagke, C., Kettermann, M., Bitsch, N., Bücken, D., Weismüller, C., Urai, J.L., 2019. The effect of obliquity of slip in normal faults on distribution of open fractures. *Frontiers in Earth Science* 7, 18.
- Walsh, J., Watterson, J., 1993. Fractal analysis of fracture patterns using the standard box-counting technique: valid and invalid methodologies. *Journal of structural Geology* 15, 1509–1512.
- Wang, Y., Jing, H., Su, H., Xie, J., 2017. Effect of a fault fracture zone on the stability of tunnel-surrounding rock. *International Journal of Geomechanics* 17, 04016135.
- Watkins, H., Bond, C.E., Healy, D., Butler, R.W., 2015. Appraisal of fracture sampling methods and a new workflow to characterise heterogeneous fracture networks at outcrop. *Journal of Structural Geology* 72, 67–82.
- Wilson, T.H., 2000. Some distinctions between self-similar and self-affine estimates of fractal dimension with case history. *Mathematical geology* 32, 319–335.
- Wilson, T.H., 2001. Scale transitions in fracture and active fault networks. *Mathematical Geology* 33, 591–613.
- Zeeb, C., Gomez-Rivas, E., Bons, P.D., Blum, P., 2013. Evaluation of sampling methods for fracture network characterization using outcrops. *AAPG bulletin* 97, 1545–1566.

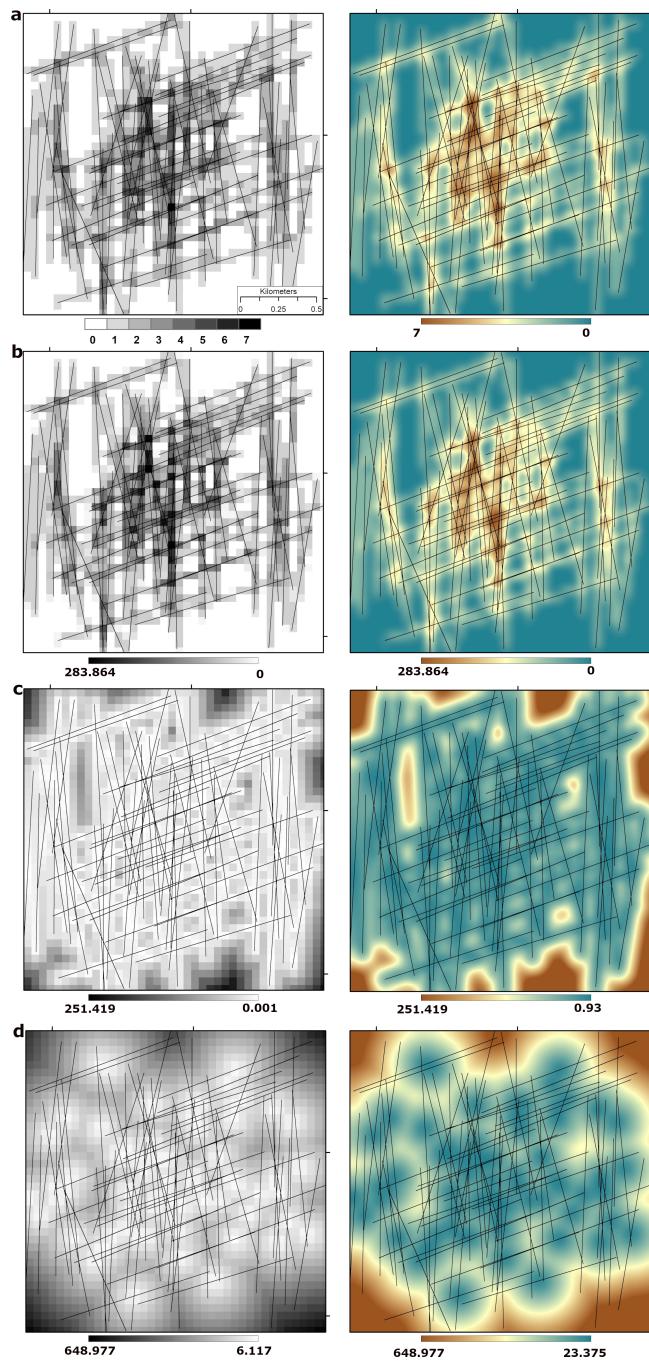


Figure 6: Examples of spatial analysis performed on a synthetic dataset comprising 75 lines with two dominant orientations. The greyscale images represent the results of rectangular window sampling and the coloured maps show the results after up-sampled by one order of magnitude. **a** density map showing intersections per pixel, **b** intensity displaying cumulative length of lines per pixel, **c** distance map of pixel centres to nearest line, and **d** distance map of pixel centres to nearest line centre.

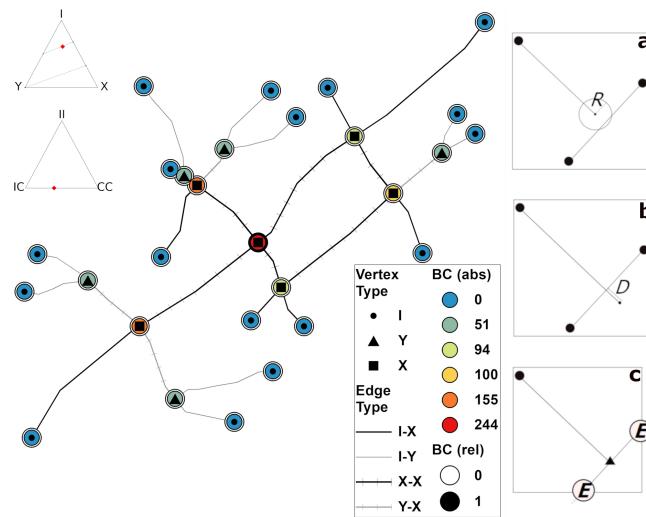


Figure 7: Example of a networks graph representation. Vertex types are labelled according to the number of outgoing edges (I:1, Y:3, and X:4). The edge types classification is based on the vertices adjacent to each edge. The colour circles surrounding every vertex show the absolute betweenness centrality (BC (abs)) and the large black and white circles show the relative betweenness centrality (BC (rel)). The two ternary plots in the upper right represent the network classification based in vertex type (I-Y-X) and based on edge connectivity (II-IC-CC). The two dotted lines in the upper ternary plot represent the average degree of vertices in a network that is related to the percolation threshold.

To the three important parameters fro building a georeferenced graph are shown schematically. **a** critical merging distance which is a buffer around each isolated vertex. If an edge is located within this buffer the edges are merged. **b** critical length governing the removal of spurs that would lead to falsely classifying a vertex as X-node instead of Y-node. **c** if a graph should be cropped to a area of interest (aoi) (i.e. the extend of a raster file) some features will be truncated along the boundaries of the aoi. To minimize the truncation bias fro statistical analysis cause by the clipping, the intersection points between line features and bounding box are labelled as edge nodes (E-nodes). Vertices labelled as E-nodes are exuded from any analysis and so are edges that contain an E-node.

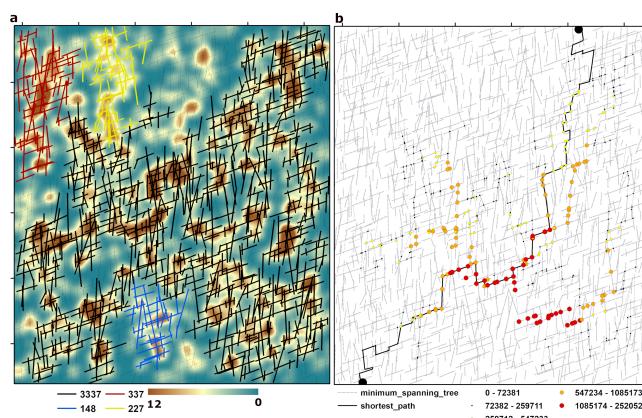


Figure 8:

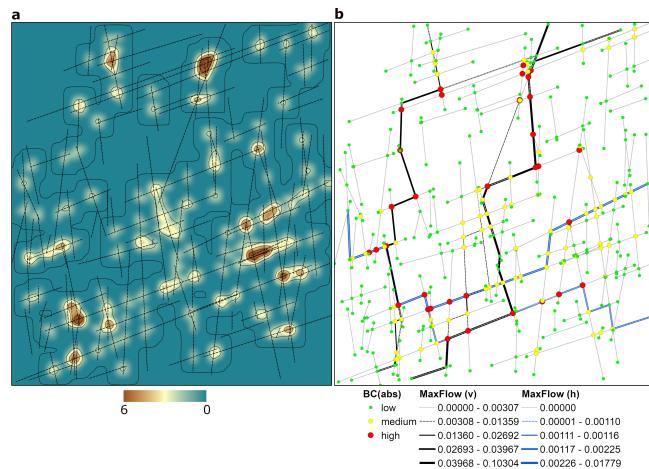


Figure 9: