Journal of Structural Geology 155 (2022) 104528

Contents lists available at ScienceDirect

Journal of Structural Geology

journal homepage: www.elsevier.com/locate/jsg

A new subsampling methodology to optimize the characterization of two-dimensional bedrock fracture networks

Nikolas Ovaskainen[a,b,*], Nicklas Nordbäck[a,b], Pietari Skyttä[b], Jon Engström[a]

[a] Geological Survey of Finland, P.O. Box 96, FI-02151, Espoo, Finland
[b] University of Turku, Department of Geography and Geology, Finland

ARTICLE INFO

Keywords:
Fracture network
Subsampling
Geostatistics
Software

ABSTRACT

This paper introduces a new subsampling method to determine the empirical relationships between the areal sampling coverage and the topological-geometric parameters resulting from characterization of two-dimensional outcrop fracture networks. We further developed these relationships into correlations between the total sampling coverage and variance of the characterization result, hence providing insight to the objective uncertainties related to fracture network characterization. The analyses were conducted using two new open-source Python packages: fractopo and fractopo-subsampling, designed for fracture network analysis and subsampling, respectively. We conducted the study on the well-exposed crystalline outcrops of Getaberget, Åland Islands, Finland, where a total of 42499 fracture traces were manually digitized from 13 circular target areas. For the purposes of subsampling, we conducted fracture network characterization for randomly located and sized subsample areas, which locate within the larger target areas.

Based on our subsampling results we provide recommendations for the preliminary optimization of areal coverage used in outcrop fracture sampling and the use of our subsampling method for assessing the precision related to the areal fracture network characterization in other previously uncharacterized areas. As an example, we recommend using a total sampling area of 8000 m^2 with 8 circular sampling areas to define the power-law exponents of fracture traces when conducting outcrop fracture network characterization with drone-based methodology done with similar initial sampling setup in comparable geological environments, as this coverage has shown an acceptable level of precision.

1. Introduction

Bedrock discontinuities such as joints and faults, here collectively termed fractures, act as pathways for fluid (hydrocarbon, water) and heat transportation in crystalline rocks (Puress and Narasimhan, 1982; Nelson, 1985) and contribute to the stability of the bedrock (Bieniawski et al., 1969). Modern remote sensing methods with Unmanned Aerial Vehicles (UAVs) enable much faster generation of large two-dimensional fracture network datasets from natural outcrops compared to the conventional field-based approaches (Bemis et al., 2014; Vasuki et al., 2014; Bisdom et al., 2017). However, extracting the fracture traces from the raster images typically still involves time-consuming manual digitizing (Vasuki et al., 2014; Palamakumbura et al., 2020), which leads to decisions being made on the spatial coverage of the mapped areas, such as the quantity and placement of potential sub-areas for fracture mapping, eventually leading to increased uncertainties within the results. Mapping uncertainties are either subjective or objective (Tannert et al., 2007). Subjective uncertainties are caused by observer bias in manual interpretation of the raster images, leading to e.g. omission of some features and overall poor reproducibility of the interpretations (Bond et al., 2007; Andrews et al., 2019; Peacock et al., 2019; Prabhakaran et al., 2019) which can be reduced by integrated field observations (Hardebol and Bertotti, 2013) or by automatic fracture mapping (Prabhakaran et al., 2019). (Semi-)automatic methods for fracture mapping have been developed (e.g. Vasuki et al., 2014; Thiele et al., 2017) but especially the fully automatic methods do not yet produce results equivalent to manual approaches (Vasuki et al., 2014; Prabhakaran et al., 2019). The objective uncertainties are caused by the constraints of nature and sampling, main issues being the availability and size of the outcrops, variable visibility of fractures on the outcrop surface caused by censoring by e.g. vegetation and limitations in sample sizes and sample area which mask the true length of fractures (Rohrbaugh

---

[*] Corresponding author. Geological Survey of Finland, P.O. Box 96, FI-02151, Espoo, Finland.
E-mail address: nikolas.ovaskainen@gtk.fi (N. Ovaskainen).

https://doi.org/10.1016/j.jsg.2022.104528
Received 24 June 2021; Received in revised form 12 January 2022; Accepted 16 January 2022

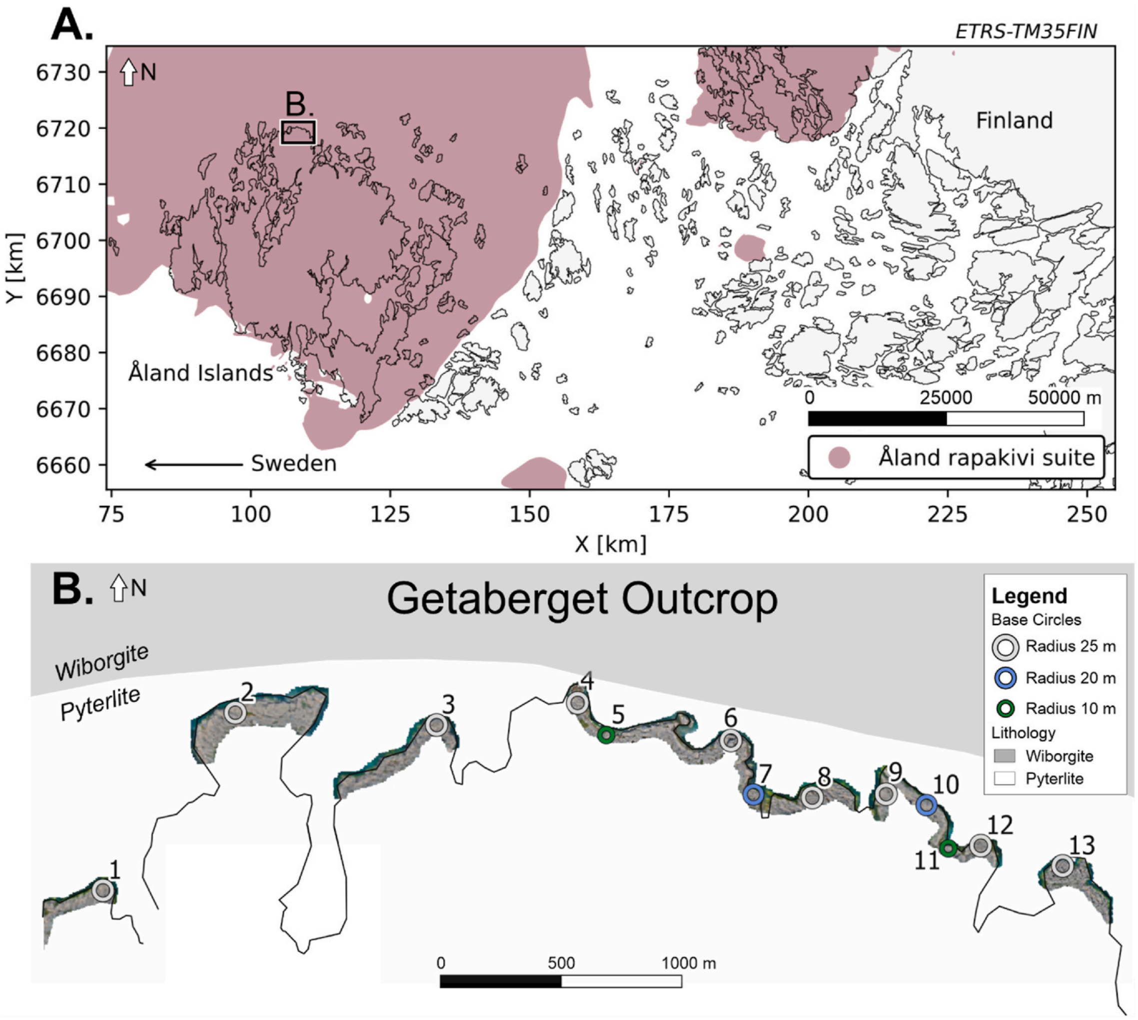

Fig. 1. A: Location of the Getaberget study site and the overall lithological setting. Location of Getaberget (subfigure B.) is indicated by the black rectangle which is located on the northern shoreline of Åland Islands, Finland. Lithology is from Bedrock of Finland 1:200 000 by Geological Survey of Finland (GTK, 2017). B: Areal lithology, processed orthomosaics and base circles. Base circles are labeled with unique IDs and the circles are color-coded based on their radii. The black polyline is the digitized shoreline. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

et al., 2002; Ortega et al., 2006; Weiss, 2008; Zeeb et al., 2013). Furthermore, fracture patterns vary spatially in the bedrock due to heterogenic bedrock conditions during fracturing. Such heterogeneities are typically caused by local variations in the mechanical properties of the host rock and by the structure-controlled local stress perturbations of the bedrock (Nickelsen and Hough, 1967; Rawnsley et al., 1992; Kim et al., 2004; Skyttä et al., 2021). For these reasons, the representativeness of the available outcrop is always uncertain. Collectively the objective uncertainties define a domain of random inputs where individual uncertainties (e.g. spatial variance, sample size) are difficult to distinguish in the final output result.

Studies on the effect of sampling on fracture network characterization involve both the overall method development and evaluation of the areal sampling coverage. Zeeb et al. (2013) used Artificial Fracture Network simulation and automatic sampling to evaluate the established fracture sampling methods (linear scan-line by Priest and Hudson (1981), circular scan-line by Mauldon et al. (2001) and areal sampling by Pahl (1981) and Priest (1993)) regarding the objective sampling uncertainty and concluded that areal sampling produces the most accurate results for most network characteristics. An exception was fracture intensity for which the circular scan-line subsampling method (Mauldon et al., 2001) was more reliable. Other previous sampling studies have given suggestions on the minimum number of fracture traces *per* target area for accurate determination of fracture characteristics in areal sampling: suggested value ranges include 150 to 300 fractures by Priest (1993), and minimum values of either 200 or 110 have been suggested by other authors (Bonnet et al., 2001; Zeeb et al., 2013, respectively). As the censoring effect caused by the target area boundary on observed fracture lengths drastically increases the uncertainty in parameter acquisition (Priest, 1993; Zeeb et al., 2013), Priest (1993) further suggested that 50% of the fractures should have at least one termination visible in the target area. For the node counting method introduced by Mauldon et al. (2001), ten circular scan-lines, with a dimension larger than the average fracture block size or spacing and containing at least 30 fracture endpoints, are required within the sample region to reach a precision of 15% for results (Rohrbaugh et al., 2002). Despite the method development and available recommendations for representative sampling by the previous studies, the direct effect of areal sampling coverage on fracture network characterization precision and accuracy has not been studied (Zeeb et al., 2013). Moreover, the previous studies have not included topological characteristics of fracture networks (Manzocchi, 2002; Sanderson and Nixon, 2015) in examinations of representative sampling.

In this paper we argue that optimization of the initial areal sampling coverage is possible through understanding the dependency of the end result (=fracture network characterization parameters) on the spatial

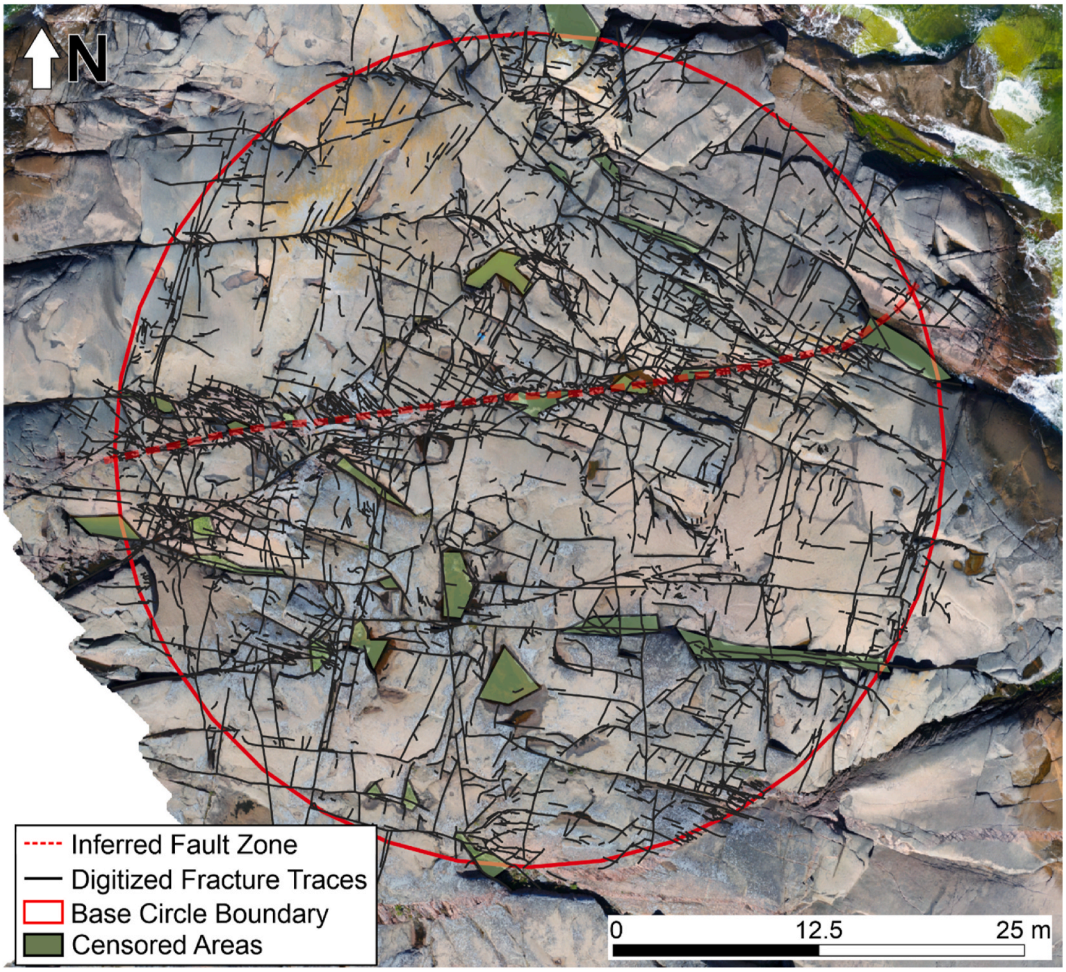

Fig. 2. Base circle 4. The censoring caused by e.g. vegetation and deep shadows are digitized as polygons within each base circle. There is a total of 3752 digitized fractures traces within this circle. We have visualized an E–W-trending Inferred Fault Zone as an example in the base circle to highlight one of many potential causes for fracture pattern variation in the Getaberget outcrop.

sampling coverage used for the characterization. We describe empirical relationships between the spatial coverage and the variance and representativeness of our dataset by randomly subsampling our extensive initial dataset and then comparing the subsampling results to the results gained from the initial dataset. Using these relationships, we justify the optimization of the sampling area and demonstrate that sufficient precision and accuracy of results is achievable with a dataset which is spatially and quantitatively more limited but still suited for the research target. Consequently, we develop the findings into preliminary recommendations about the required areal coverage for fracture mapping, and this will lead to saving time and effort used for the manual generation of fracture datasets when applied to new comparable mapping targets. In addition, subsampling can be used to determine the general variance of characterization results with different configurations of data.

Our primary dataset consists of 42499 fracture traces which were manually digitized from drone-acquired outcrop orthophotographs using areal sampling comprising of 13 initial circular target areas (=base circles; Fig. 1). The dataset was collected from shoreline outcrops at Getaberget, Åland Islands, Finland. In addition to the primary dataset, we use a secondary dataset of field mapped faults to evaluate the effects of faults and their proximity on the base circle fracture network characteristics. To conduct the subsampling, we place randomly located and sized subsample circles within the base circles and characterize both the base circles and subsample circles using geometric and topological fracture network parameters. Consequently, we compare the characterization results from the subsample circles to characterization results from the base circles, analyze the revealed correlations between coverage and characterization results and discuss the underlying uncertainties that affect the characterization.

2. Geologic setting

The site of this study is formed by a roughly E–W trending 4 km long chain of well-exposed, glacially polished, semi-continuous outcrops at Getaberget, Åland islands, Finland (Fig. 1). Lithologically the outcrops consist of homogeneous pyterlitic rapakivi granite (Fig. 1) which was emplaced 1.58–1.57 Ga ago (Laitakari et al., 1996) during a Mesoproterozoic anorogenic event (Luosto et al., 1990; Haapala and Rämö, 1992; Rämö and Haapala, 2005) associated with N–S crustal extension (Nironen, 1997). In the immediate proximity of the outcrops is the contact between Pyterlite and Wiborgite with a roughly WNW–ESE strike (Fig. 1; GTK, 2017). Since the rapakivi granite is of Mesoproterozoic origin it has not been subjected to penetrative ductile tectonic overprint and therefore lacks metamorphic precursor fabrics or larger structures which are characteristic in Precambrian crystalline rocks elsewhere in Finland (Nironen, 1997). Overall, these factors make the location ideal for two-dimensional fracture network characterization as there is no need to consider the likely effects of ductile precursor fabrics, larger ductile structures (e.g. shear zones) or significant lithological variation on the overall fracture patterns (See also Skyttä et al., 2021) or on the selection of the sampling sites. No prior knowledge exists about the brittle structures in the area and therefore we have performed

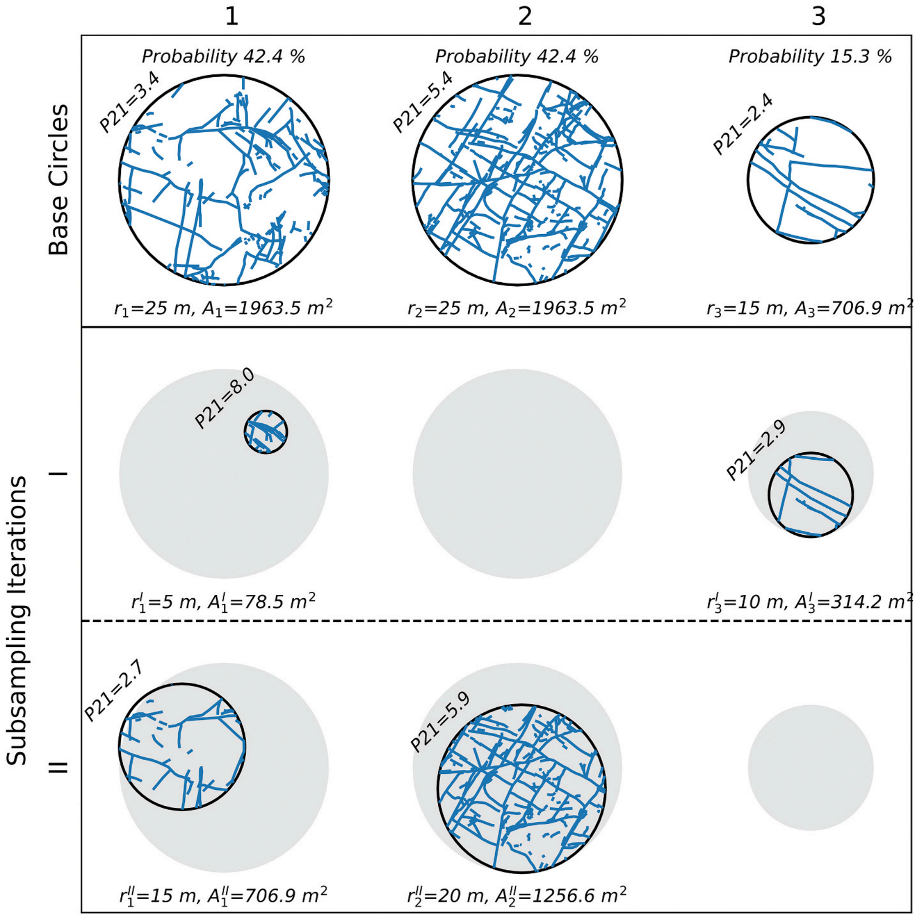

Fig. 3. Visualization of the subsampling workflow. The topmost circles represent three base circles (1, 2, and 3) from which randomized subsampling is performed. The next rows represent two subsampling iterations (Iterations I and II) where a random amount of the base circles is chosen. Inside the chosen circles sample circles of random radii (*r*) are placed and subsequently characterized. For subsampling, the base circle selection probability is weighted by the area (*A*) of the base circles. Example probabilities for the selection are displayed above the base circles. Individual circle characterization values are shown for Fracture Intensity P21 (Dershowitz and Herda, 1992; Sanderson and Nixon, 2015).

structural mapping of brittle fault structures occurring within the outcrops (Chapter 4.3) and consider their effect on the fracture pattern in our discussion.

3. Methods

3.1. Data acquisition

We acquired aerial photographs spanning most of the well exposed outcrops at Getaberget (Figs. 1 and 2) using an Unmanned Aerial Vehicle (=UAV; DJI Phantom 4 Pro). UAV flights were planned with the freely available Pix4Dcapture autopilot software. All photographs were taken with a flight elevation of 20 *m* and with a minimum overlap of 60–70% between the images. The resulting ground sampling distance is 0.55 *cm*/*pixel*. Agisoft Metashape software (Agisoft, 2020) was used to process the photographs into accurately georeferenced orthomosaics with 10 VRS-GPS ground control points *per* UAV flight. Our UAV surveys and photogrammetric processing follows the workflow described by James (2017). In total the photographs covered an area of roughly 328 000 m^2 along the Getaberget shoreline of which 20 800 m^2 was digitized for this paper.

We placed 13 circular initial base target areas for fracture digitizing along the continuous parts of the Getaberget shoreline outcrops in roughly 200 ± 50 *m* intervals. Local gaps in exposure of up to 500 *m* extended the interval locally and the placement was skewed towards better exposed parts of the outcrops (Fig. 1). The purpose of the interval sampling was to avoid the bias of placing target areas based on any previous geological interpretation of fracture patterns and locations of potential faults. E.g., base circle 4 was placed without consideration of the inferred E–W-trending intersecting fault zone (Fig. 2). To avoid orientation bias in node counting (Mauldon et al., 2001; Watkins et al., 2015) a circular target area shape was chosen. Because the outcrop axis perpendicular to the shoreline locally constrained the local maximum base circle size, the circle size varies with a maximum radius of 25 *m* and a minimum radius of 10 *m* (Fig. 2). Within each target area the extent of censoring that limits fracture trace digitizing (e.g. dark shadows, vegetation) was digitized as polygons into a single database file (Fig. 2).

We digitized fracture traces from the drone orthomosaics using ArcGIS 10.6.1 and QGIS 3.14 (QGIS Development Team, 2020) and used the built-in snapping functionality within the software to document the topological abutting and cross-cutting relationships between the fractures (Fig. 2) following recommendations and workflow by Nyberg et al. (2018). During digitization the verification of fracture type (e.g. joint, vein, fault, etc.) is not possible due to the polished nature of the horizontal outcrop surfaces and the limited resolution of the orthophotographs and therefore we could not discriminate between types. Digitizing was done by two operators at the Geological Survey of Finland. We validate the digitized traces using the fractopo validation tools to eliminate conflicting topological intersections between traces such as V-nodes (Sanderson and Nixon, 2015) and other ambiguity in trace interactions (See fractopo *documentation*). We further use QGIS 3.14 (QGIS Development Team, 2020) and geopandas (Jordahl et al., 2020) to visualize the digitized traces (See e.g. Fig. 2).

3.2. Fracture network subsampling

To address the effect of spatially varying fracture sampling areas, we introduce a method that subsamples the fracture network with randomly located and sized samples of the network (Fig. 3). Because the subsampled fracture network traces are geometries, we cannot simplify

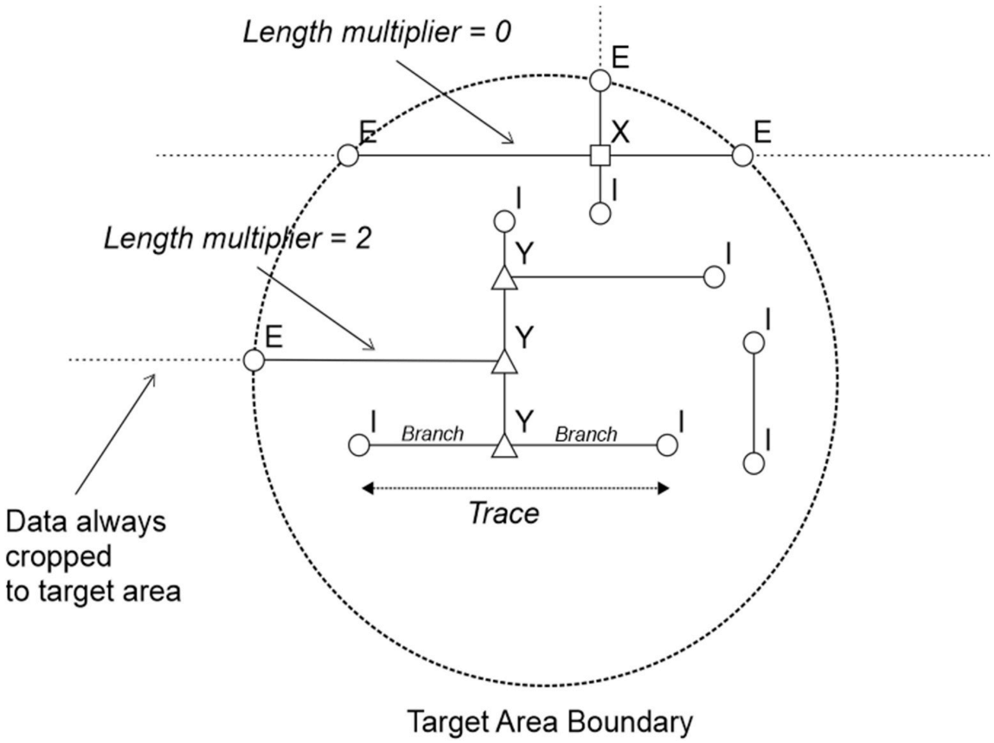

Fig. 4. Diagram illustrating the topological relationships (nodes X, Y, and I) between digitized traces and the assigned boundary length multipliers. Traces are the digitized continuous lines ending inside the circle in either I- or Y-nodes or outside the circle boundary, in which case the trace is during analysis cropped to the circle boundary, resulting in an E-node endpoint at the boundary. Branches are the trace segments between two nodes of any type (Sanderson and Nixon, 2015). The boundary length multipliers are applied using the same rationale to individual traces and branches.

them to scalar input variables and therefore we cannot model input variable distributions. Thus, the subsampling does not fit the criteria for Monte Carlo simulation (Mooney, 1997; Raychaudhuri, 2008).

We perform the randomized fracture network subsampling in two stages.

1. Within each digitized base sample circle, random circles varying in both radius and location are placed and fracture network attributes are collected.
2. A random amount of the base circles is chosen, with selection probability weighted by the base circle area. For each chosen base circle a single random circle from the first stage is chosen with a random radius and location (Fig. 3). Multiple subsamples are not taken from the same base circle in the same iteration.

Both stages are iterative. The first stage is conducted multiple times and collected into a dataset where each row represents a single subsample circle within one of the base circles. The subsample circle radius, x and y coordinates are randomly sampled from a uniform distribution (resulting in non-uniform area values of subsample circles as circle area has a polynomial correlation with its radius.). In the second stage we use the collected first stage data and again iterate multiple times to form the final subsampling dataset where each row represents an overall network characterization consisting of aggregated results from multiple subsample circles, equivalent in methodology to the initial full base circle characterization, only with data of different configurations varying both in location and sampling coverage.

Fractures are subsampled in the first stage by cropping the base circle fracture traces to the subsample circle boundary (Fig. 3). In the first stage the maximum random radius is the radius of the base digitized circle (Figs. 2 and 3) and the mean value of fracture block size places a limit on the minimum circle size for assessment of network properties (Rohrbaugh et al., 2002). Therefore, we use a minimum radius of 5 *m* for subsample circles based on inspection of the fracture trace data to completely rule out the possibility that any subsample circle is smaller than the fracture block size. We use multiple different filters to avoid previously noted uncertainties related to low trace counts and areal censoring (e.g. shadows, vegetation) before passing the first stage dataset to the second stage. The relative amounts of areal censoring within the subsampled circles are determined using the digitized censoring polygons (Fig. 2) and subsampled circles with relative censoring of over 15% of the sample area are removed. Furthermore, circles with less than 110 traces or with less than 30 trace endpoints (I- and Y-nodes combined) are removed according to recommendations by Zeeb et al. (2013) and Rohrbaugh et al. (2002), respectively.

We address the varying base circle radius (10–25 *m*; Fig. 1) by weighting the base circles with their total area values in the second stage so that base circles with larger radii are more likely to be picked than smaller base circles (Fig. 3). The weighting allows the inclusion of smaller base circles in the subsampling without them skewing the subsampling results. Subsequently, larger circles have more weight on the aggregated averages. This is emphasized further as we aggregate results in the second subsampling stage with an area-weighted average. E.g. for subsampling iteration I in Fig. 3 the aggregated *P21* value would be:

$$\frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{8.0\frac{m}{m^2}\times 78.5m^2 + 2.9\frac{m}{m^2}\times 314.2m^2}{78.5m^2 + 314.2m^2} = 3.92\frac{m}{m^2}.$$

After the second stage a dataset of 100 000 subsampling iteration results is collected. Each row in the dataset represents a single subsampling iteration in which 1 to 13 base circles have each been subsampled and the results from the subsample circles aggregated. Of primary interest for result interpretation are the subsample circle count (varying between 1 and 13) and total subsample area (maximum of 20080 m^2 which is the total area of the base circles) variables as they together represent the spatial coverage of each iteration. We group the second stage dataset using these two variables by first choosing appropriate value ranges for both variables and then pairing all value ranges of both groups with each other into every possible group pair combination. These represent different configurations of sampling coverage. E.g. a low circle count and low total area pair represents very low coverage, or a high circle count and high total area pair represents high coverage. We visualize the distribution of the selected fracture network parameter values for each group pair with seaborn boxplots (Waskom, 2021) and display group pair iteration counts in a table. We conduct the described full subsampling workflow of running parallel random network subsampling and result aggregation using fractopo-subsampling, a collection of Python scripts and notebooks, that wraps the functionality of fractopo (See Section 3.3.). When interpreting results from subsampling we use the term *accuracy* to describe the tendency of the median of repeated subsamples to describe the full dataset by comparing the median with the full dataset reference values and *precision* to describe the scatter of individual results around the median result.

3.3. Fracture network characterization

To conduct the fracture network characterization, we present fractopo, an open-source Python package developed for two-dimensional fracture and lineament trace map analysis inspired by FracPaQ (Healy et al., 2017) and NetworkGT (Nyberg et al., 2018). FracPaQ, among many other features, implements the circular scan-line subsampling method to reliably determine fracture intensity and NetworkGT first implemented a methodology to determine topological branches and nodes (Fig. 4) from fracture trace data. fractopo enables the complete fracture network characterization fully within Python without the need for GIS-software or any proprietary software, such as matlab, which in turn allows it to be more readily useable as a library rather than only as a toolkit or plugin within a single software environment. Usage is possible both from a Python interface or alternatively from the command-line for e.g. trace validation. It determines topological branches and nodes and uses them along with the traces to determine and visualize a range of two-dimensional fracture network characteristics similarly to NetworkGT (Nyberg et al., 2018). We further use free and open-source software to visualize different types of data related to network characterization: geopandas (Jordahl et al., 2020) for spatial data, seaborn (Waskom, 2021) for statistical data and QGIS 3.14 (QGIS Development Team, 2020) for general maps.

We conduct fracture network characterization with the fractopo package for both the 13 base circles and all random subsample circles. The fractopo package creates random network subsamples with fully determined characteristics from given fracture traces and circular base circle where each random subsample is a smaller random radius circle fully within the base circle. For each target circle, base or subsampled, the digitized traces are cropped to the circle boundary and thereby all analysis results are only determined inside the circle (Fig. 4). We use topological node counting to determine network parameters to avoid uncertainty related to traces that leave the target area, following the recommendations of many recent studies (e.g. Procter and Sanderson, 2018; Nixon et al., 2020; Alvarez et al., 2021; Hansberry et al., 2021). Our characterization workflow has similarities to recent workflows (Bertrand et al., 2015; Bisdom et al., 2017; Loza Espejel et al., 2020; Ovaskainen, 2020; Skyttä et al., 2021).

We characterize both base and subsample circles with four geometric and topological fracture network parameters:

- *Fracture Intensity P21*, based on total fracture length in area (Dershowitz and Herda, 1992; Sanderson and Nixon, 2015).
- *Dimensionless Intensity B22*, based on total fracture length divided by area and multiplied by mean branch length (Sanderson and Nixon, 2015).
- *Connections per Branch* (Sanderson and Nixon, 2015).
- Fitted *Power-law Exponent* for fracture trace distribution, determined with powerlaw Python package (Clauset et al., 2009; Alstott et al., 2014).

P_{21} fracture intensity is the mean value of trace length $L_t[m]$ *per* unit area $A\ [m^2]$ following Sanderson and Nixon (2015):

$$P_{21} = \frac{\sum L_t}{A}$$

B_{22} dimensionless intensity is calculated by dividing the total branch length within an area by the area value and then multiplied by the mean branch length $B_C\ [m]$ following Sanderson and Nixon (2015):

$$B_{22} = \frac{\sum L_b}{A} \times B_C$$

Connections per Branch (C_B) is used to quantify the connectivity of a fracture network by counting topological nodes (X, Y, and I; N_X, N_Y, and N_I, respectively) of a network (Sanderson and Nixon, 2015). The node counts are first used to calculate the quantity of branches N_b with:

$$N_B = \frac{1}{2}(4N_X + 3N_Y + N_I)$$

Consequently, *Connections per Branch* can be derived by replacing N_b with the above:

$$C_B = \frac{3N_Y + 4N_X}{N_b}$$

$$C_B = \frac{3N_Y + 4N_X}{\frac{1}{2}(4N_X + 3N_Y + N_I)}$$

Following the study by Zeeb et al. (2013) we describe the fracture trace length distribution with a power-law fit (Pickering et al., 1995; Odling, 1997; Bonnet et al., 2001). A fractal distribution $n(l)$ of trace lengths l can be described with a power-law exponent a and a constant A (Bonnet et al., 2001):

$$n(l) = A \times l^a$$

We display a values as negative in results and plots. The exponent is derived in code from powerlaw.Fit alpha α attribute (Alstott et al., 2014) by subtracting 1 and negating the value ($a = -(\alpha - 1)$). The powerlaw package automatically determines the power-law cut-off value which is the lowest fracture trace length fitted to the power-law trend. The need for the cut-off is typically attributed to the effect of the scale of observation and consequent inability to sample the smallest fractures accurately (Pickering et al., 1995; Bonnet et al., 2001). We display fracture trace length fits, in addition to power-law, for exponential and lognormal models (See e.g. Pickering et al., 1995; Bonnet et al., 2001) for the fracture trace length distributions of the 13 base circles. We compare lognormal, exponential and power-law fits to the length data to estimate the accuracy of the power-law fit in modelling the length distribution as recommended by Clauset et al. (2009). We compare the fits using loglikelihood values R and p (Clauset et al., 2009; Alstott et al., 2014). The loglikelihood ratio R is negative when the lognormal or exponential fit is more likely a better fit and positive when power-law is more likely as the better fit. The significance of the ratio is given by the p value. When the value of p is below 0.1 we interpret the ratio as non-significant and when above we interpret it as significant (Clauset et al., 2009). If we can accept the hypothesis that power-law fits can model the length distributions of the base circle traces based on the displayed comparisons to lognormal and exponential we will only consider the fractal, single power-law, approach in our length distribution characterization conducted with subsampling. We fit the *Power-Law Exponent* regardless of the goodness-of-fit of other possible distribution models (e.g. lognormal, exponential) as it is widely used as a fracture network characterization parameter in crystalline rocks (e.g. Follin et al., 2014; Bertrand et al., 2015), because there is physical rationale for the cause for fracture properties such as length to follow power-law models (Bonnet et al., 2001) and more importantly because in this paper we aim to introduce a methodology with only a first-pass assessment of the Getaberget outcrop fracture data. If required, the introduced subsampling methodology can be readily used for exploring also exponential or lognormal distribution fits to trace length subsample data (fit parameters for all three distributions are included in the Stage 1 dataset) as subsampling is agnostic in relation to the parameter that is calculated from the subsamples.

Because fractures are cropped to the boundary of target circles the full fracture length is censored (Fig. 4). This censoring effect is inherent in all fracture sampling (Mauldon et al., 2001). We address the effect of fracture length censoring by giving weights to fracture lengths based on fracture intersection with the target circle boundary (Fig. 4). We multiply the length of all fracture traces or branches that do not intersect the boundary by one, the lengths of fractures that intersect the boundary edge once by two and reduce the length of fractures that intersect the edge twice to zero based on the following reasoning: Length of fractures fully within the circle is precisely defined, however that is not the case for fractures that intersect the boundary once or twice. When considering a homogeneously distributed fracture population there is an equal probability that the centers of fractures that intersect the circular boundary once have their centers inside or outside the boundary (Mauldon, 1998; Zeeb et al., 2013). Thus, with a large enough fracture population the average fracture center is at the boundary and only half of the average fracture length is inside the circle. Consequently, we claim that the length of such fractures can be included in further analysis with a multiplier of two. However, no control on fracture length is possible for fracture traces or branches that intersect the boundary twice and therefore the unknown length is reduced to zero, thereby discarding such traces or branches from analysis. This length-weighting is applied for all analyses of fracture trace and branch lengths, such as length distribution modelling and length-weighted orientation analysis where the length of individual fractures must be defined. The only exception is that for fracture intensity determination the full fracture length strictly inside the circle is used. The aforementioned method to control the area boundary censoring will result in censoring of fractures traces that are longer than the circle diameter by design which should be noted when analyzing fracture trace length distributions. Traces longer than 20 *m* (diameter of smallest base circle) are rare in the Getaberget outcrops and their continuation all the way through the circles is highly subjective but for subsample circles, which have a minimum diameter of 10 *m*, such traces will occur. Furthermore, due to the limited resolution of the orthophotographs small fractures will also be censored which has implications for results of *Fracture Intensity P21* and *Trace Power-law Exponent*. *Fracture Intensity P21* is a scale-dependent parameter and therefore the effect is expected (Ortega et al., 2006; Sanderson and Nixon, 2015) but for *Trace Power-law Exponent* the censoring requires the use of a cut-off to fit the power-law trend (Pickering et al., 1995; Bonnet et al., 2001).

3.4. Base circle characterization and reference values

In addition to the four previously introduced geometric and topological fracture network characterization parameters, we perform a more rigorous characterization of only the 13 base circles to better inspect the base circle network characteristics and to display the variation between them. We plot fracture trace length distribution fits for power-law, lognormal and exponential to estimate the accuracy of the power-law model in describing the distribution by comparing the fits with each other (Clauset et al., 2009; Alstott et al., 2014). The fitting is done to tail-truncated length data for all three fits even though only power-law requires a cut-off. This is done to allow meaningful comparisons between the fits which would not be possible if the fits are assigned to datasets of different sizes (Clauset et al., 2009; Alstott et al., 2014). We visualize fracture trace orientation distributions of all base circles with length-weighted equal-area rose plots (Sanderson and Peacock, 2020) to assess the variation of base circle fracture orientations. To assess the spatial variability of the previously selected four fracture network parameters (See Section 3.3) we plot each fracture network parameter value as a function of the x-coordinate of the base circle centroid and fit a linear regression trend to the values. We use seaborn (Waskom, 2021) for the regression fitting and in addition to the fit we display the confidence intervals of the fit. In essence, we project the two-dimensional base circle centroids into a one-dimensional E–W-oriented x-axis along the Getaberget study area (Fig. 1). This simplifies the determination of potential horizontal trends but disregards any possible vertical trends in the visualization. We interpret the resulting plot as such: there is no apparent spatial E–W-trend for a network parameter if the linear regression fit is horizontal while taking into account the confidence intervals of the fit. If a clear trend is observable for a parameter, it can be expected that when subsampling for that parameter the variation is partly caused by the observed trend rather than other sources of objective uncertainty.

For use as reference values in subsampling, we aggregate the four characterization parameter values of the 13 base circles by conducting an area-weighted average for the selected four characterization parameters. We use area-weighting to emphasize the results from larger circles compared to smaller circles. For the base circles, the aggregated reference values represent the result of our workflow to two-dimensionally characterize the overall Getaberget outcrop fracture network and comparisons with the aggregated subsampling results are

Table 1
Base circle characterization results. Columns are the base circle IDs as laid out in Fig. 1. Units are indicated when applicable.

	1	2	3	4	5	6	7	8	9	10	11	12	13
Diameter [m]	50	50	50	50	20	50	40	50	50	40	20	50	50
Number of Traces [a]	1856	4465	8419	3605	550	7319	1622	2225	2953	2249	358	1739	3291
Number of Branches [a]	4264	10215	19588	7408	1149	18631	3511	4546	6605	4843	590	3712	7391
Trace Min Length [m] [b]	0.05	0.04	0.04	0.02	0.03	0.05	0.04	0.04	0.04	0.04	0.05	0.01	0.05
Trace Max Length [m] [b]	31.78	29.28	27.44	21.56	8.41	32.05	20.11	26.97	34.86	19.9	14.6	25.38	31.83
Trace Mean Length [m]	1.5	0.86	0.77	0.94	0.98	0.83	1.26	1.35	1.18	0.95	1.0	1.53	1.15
Branch Mean Length [m]	0.67	0.38	0.34	0.47	0.5	0.33	0.6	0.69	0.54	0.45	0.64	0.74	0.53
X	507	1259	2299	638	94	2223	272	372	674	441	37	318	635
Y	1394	3232	6571	2527	411	6866	1345	1577	2304	1712	158	1337	2830
I	2319	5699	10267	4683	690	7772	1899	2874	3603	2786	558	2141	3753
E	111	143	255	132	63	228	118	159	143	116	40	118	151
$C-C$	2333	5440	10951	3600	578	11923	1939	2210	3529	2562	174	1961	4273
$C-I$	1457	3746	6814	2838	411	5464	1141	1686	2437	1690	251	1275	2361
$I-I$	420	958	1696	906	129	1130	372	571	568	533	145	417	682
Fracture Intensity P21 $[m/m_2]$ [b]	1.47	2.0	3.38	1.77	1.83	3.17	1.69	1.6	1.83	1.75	1.21	1.41	1.99
Dimensionless Intensity B22	0.99	0.77	1.14	0.82	0.91	1.05	1.02	1.1	0.99	0.79	0.77	1.04	1.05
Connections *per* Branch	1.46	1.44	1.48	1.37	1.4	1.58	1.46	1.37	1.45	1.42	1.05	1.42	1.49
Trace Power-law Exponent	−2.02	−2.05	−2.3	−2.2	−1.67	−1.64	−1.37	−1.39	−2.08	−1.49	−1.67	−1.24	−1.33
Trace Power-law Cut-off [m]	5.11	2.32	3.49	4.39	1.66	1.27	1.74	1.59	4.0	1.45	1.22	1.33	1.19
Trace Power-law vs. Lognormal R	−0.12	−1.05	−0.55	−0.55	−0.68	−2.06	−2.05	−2.94	−1.08	−1.56	−0.62	−3.42	−2.82
Trace Power-law vs. Lognormal p	0.903	0.296	0.584	0.584	0.494	0.039	0.04	0.003	0.28	0.118	0.537	0.001	0.005

[a] Determined from node counting.

[b] Boundary weighting as described in Section 3.3 is not applied for these parameters.

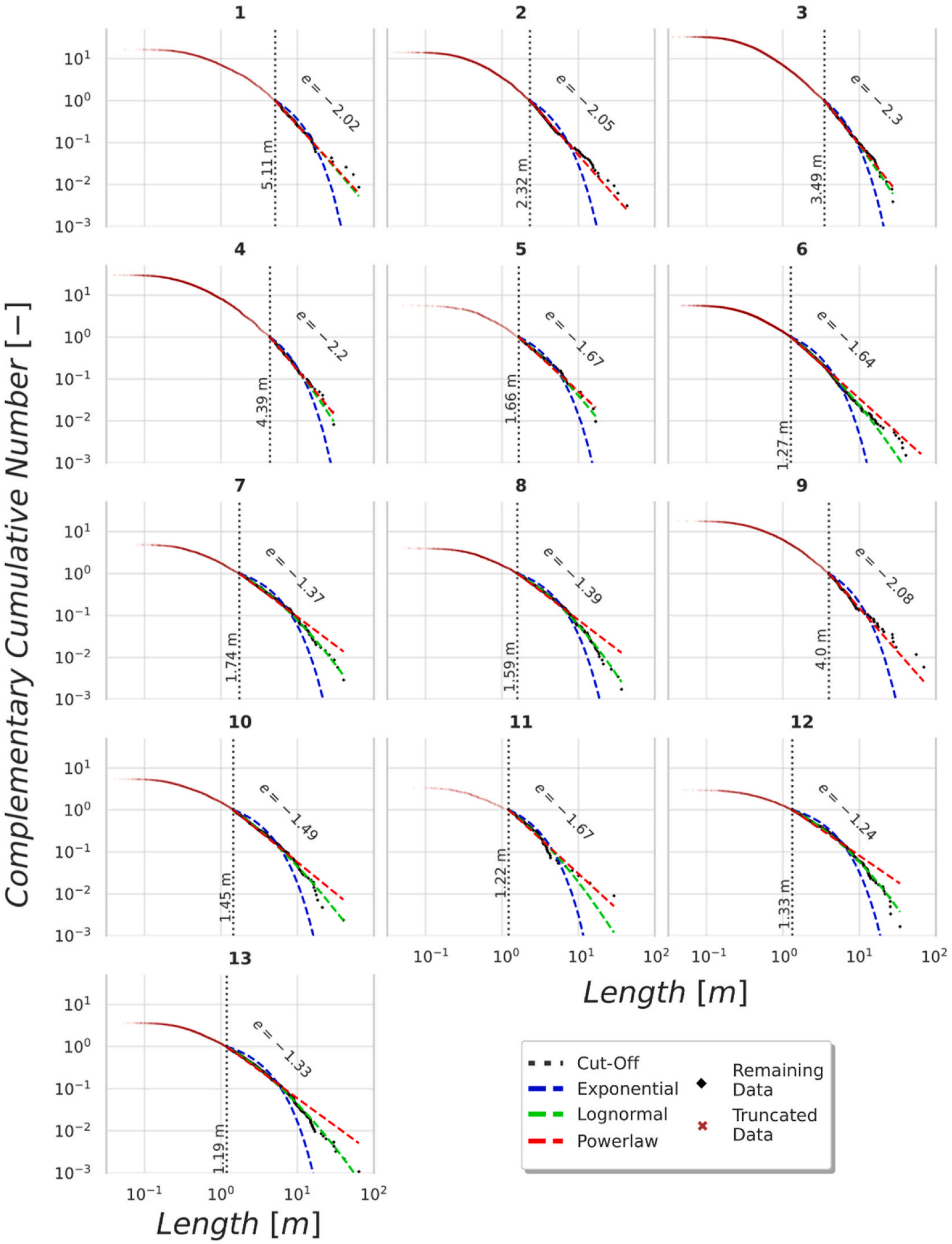

Fig. 5. Exponential, lognormal and power-law distribution fits to fracture trace length data of all base circles. Automatically determined power-law cut-offs are indicated with dotted vertical lines and the cut-off value is displayed next to the lines. Exponent of the power-law fit is also displayed (e = exponent).

used to assess if the same results could have been achieved with less sampling and to assess the effect of spatial coverage on error margins.

3.5. Characterization of faults and fault zones

As all the base circles are located within the pyterlite (Fig. 1B), we expect negligent variation caused by lithological variation or ductile features. However, potential systemic fracture pattern perturbations may be related to distinct brittle deformation zones, and these need to be evaluated to determine possible causes for the variation we observe in the area and whether if it is justifiable to aggregate data from the whole area when summarizing subsampling results. For this purpose, we mapped potential faults and fault zones in the field by recognizing fault-related secondary fracturing (e.g. horse-tails and step-overs) and damage zones (Kim et al., 2004) as displacement markers or slip lineations were rarely available. We defined fault kinematics and measured geometric properties such as primary fault plane orientation and estimated fault lengths. Faults included in the characterization had a minimum interpreted length of 15 m and they comprised recognizable secondary fracturing. The observed fault lengths are only approximations as the lengths are censored by the shoreline, vegetation and quaternary deposits. Further, due to the dominantly horizontal outcrop surfaces we could not map subhorizontal faults or normal/reverse faults with lengths above 15 *m*. In Fig. 2 we display an example of a inferred fault zone that

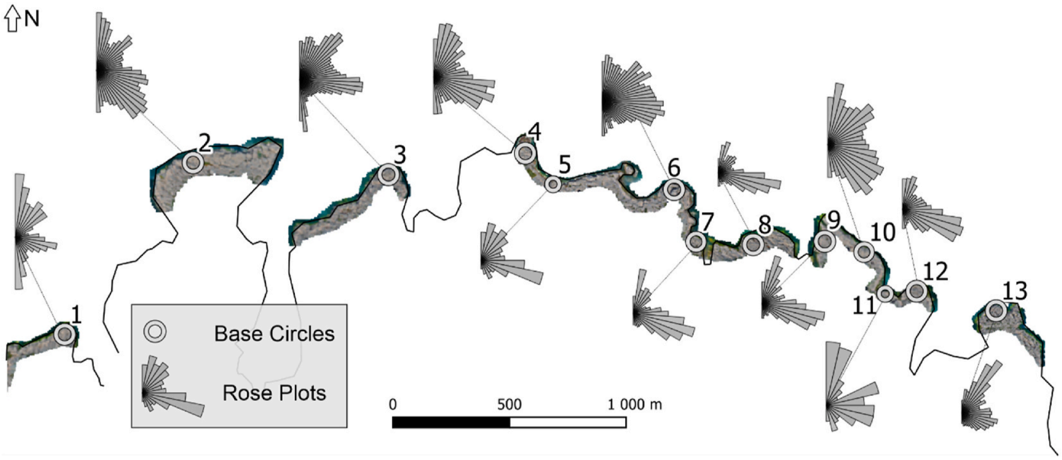

Fig. 6. Fracture trace orientation of all base circles visualized with length-weighted equal-area rose plots. The colored circles represent base circles and their radii are visualized with colors. The background rasters are the drone orthomosaics. The black continuous line is the Getaberget shoreline. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

affects the areal fracture pattern but which we could not verify as a fault during field reconnaissance. Based on our fracture network characterization results (Section 4.1) and fault and fault zone characterization (Section 4.3), in the discussion we evaluate the effect of faults in reference to the base circle fracture networks by inspecting deviations of network characterization parameters from the regionally dominant character in relation to the proximity to faults. If the mapped faults are spatially scattered without apparent systematics their controlling effect on the base circles is simply another source of objective uncertainty which is hard to distinguish from the other sources of sampling uncertainty, but if they occur systematically, their effect on the fracture pattern could be modelled (Kim et al., 2004; Choi et al., 2016; Skyttä et al., 2021).

4. Results

4.1. Characterization of base target circles and determination of reference values

All scalar base circle characterization results are collected in Table 1. Columns represent the base circle IDs and different general, geometry and topological parameters are presented on individual rows. Trace length distribution parameters are only shown for power-law fits (exponent and cut-off) but the values for the loglikelihood comparisons (Clauset et al., 2009; Alstott et al., 2014) between power-law and lognormal fits are shown to estimate the goodness-of-fit of power-law in comparison to lognormal. The exponential fit is poor for all base circles as visualized in Fig. 5 and its likelihood is therefore not examined further. The lognormal fit is more likely for all base circle trace length distributions according to the *R*-values and the *p*-value is sufficiently low ($\leq$0.1; Clauset et al., 2009) for base circles 6, 7, 8, 12 and 13 to accept the hypothesis that the lognormal fit is the most accurate. For the other base circles power-law cannot be completely ruled out as a more accurate fit in comparison to the lognormal one. High *p*-values might be caused by insufficient data which is apparent for base circle 1 where both the cut-off and *p*-value are the highest of all base circles whereas 6, 7, 8, 12 and 13 have relatively low cut-offs resulting in more data.

Spatial differences in the fracture trace orientation are visualized in Fig. 6 with length-weighted equal-area rose plots (Sanderson and Peacock, 2020). A preferred orientation, WNW-ESE, can be identified in all rose plots, although, with varying intensity. In addition, sets of N–S- and NE–SW-trending fractures occur in many of the base circles, with clear orientation maxima in plots (For N–S: 1–3, 7, 10, 11; For NE–SW: 1–3, 6, 10, 13).

Results from Table 1 for *Fracture Intensity P21, Dimensionless Intensity B22, Connections per Branch* and *Trace Power-law Exponent* are visualized in Fig. 7 to better distinguish anomalous values and potential trends. Base circle locations are plotted in Fig. 7A and characterization results on scatterplots, including fitted linear regression trend lines, in Fig. 7B–D. None of the fitted trend lines are completely horizontal but if a horizontal trend line can be fitted inside the confidence interval, we interpret that there is no apparent systematic E–W spatial trend for the parameter. A horizontal trend-line fits inside the confidence intervals of *Fracture Intensity P21, Dimensionless Intensity B22 and Connections per Branch* parameters. In contrast, there is an apparent E–W trend for *Trace Power-law Exponent* which should be kept in mind for its subsampling results. *Fracture Intensity P21* values are significantly elevated at base circles 3 and 6, highlighting the uncertainty caused by naturally varying fracture patterns of the bedrock. Based on the orientation and spatial variance visualizations (Figs. 6 and 7) we interpret that no clear trend or systematic E–W variation in the overall fracture pattern is present in the mapped area.

The base circle fracture network characterization results are aggregated in Table 2. These values are used in comparisons with the subsampling results. The *Connections per Branch* reference value of *1.4435* indicates that the overall network is more dominated by I-nodes and thus is only moderately topologically connected (Manzocchi, 2002; Sanderson and Nixon, 2015) though a more accurate network connectivity estimate depends on the proportions of X- and Y-nodes (Sanderson and Nixon, 2018) which cannot be directly interpreted from *Connections per Branch*. The *Trace Power-law Exponent* value can be compared to prior investigations of fracture length distributions of crystalline rocks (e.g. Follin et al., 2014; Bertrand et al., 2015; Ovaskainen, 2020) which tend to gravitate towards values close to *-2*. Comparisons of *Fracture Intensity P21* and *Dimensionless Intensity B22* values to earlier studies done elsewhere is mostly irrelevant due to the significant number of geological factors controlling their magnitude and spatial variance (Nelson, 1985). No fracture intensity studies are known to the authors from near the Getaberget locality.

4.2. Subsampling

We grouped the subsampling results based on subsample circle count and total area within the subsample circles (Table 3). The total area group from 8000 m^2 to 20 800 m^2 has a larger area interval because as the total area approaches the sum of total base circle area (20 800 m^2) the possible variance of subsampling results diminishes. Moreover, we are mostly interested in the results with lower total area as they

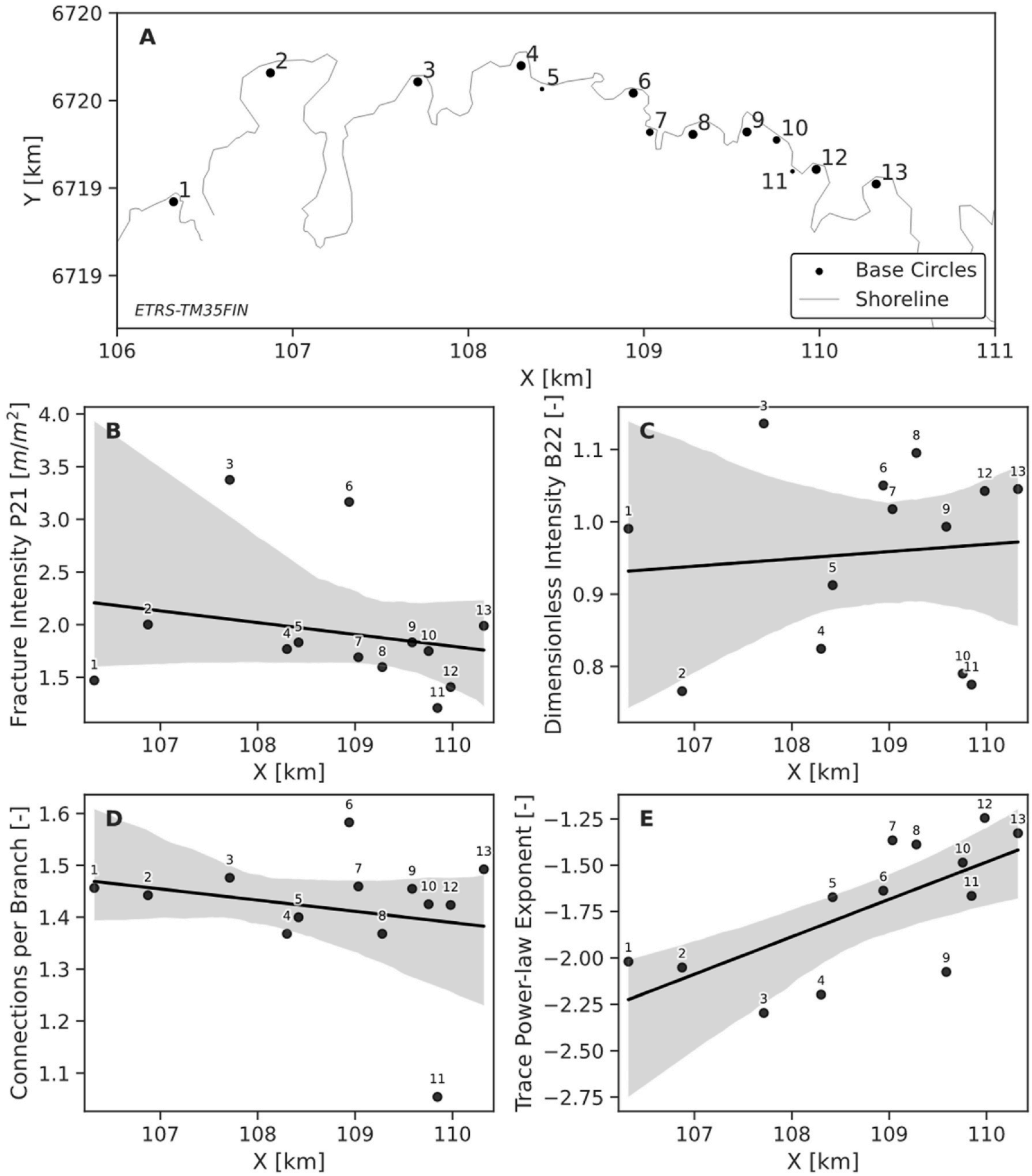

Fig. 7. A: Location of base circles. (B,C,D,E): Plots of base circle x-coordinates against their values for the selected fracture network parameters (Fracture Intensity P21, Dimensionless Intensity B22, Connections *per* Branch, Trace Power-law Exponent, respectively). A linear regression trend is fitted to the scattered values. The fit is in the shaded area (confidence interval) with a 95% confidence. A horizontal trend-line fit indicates no apparent spatial correlation in an E–W orientation.

Table 2
Area-weighted averages of network parameter values from the 13 base circles. Used as reference values when comparing with subsampling results.

Parameter	Value
Fracture Intensity P21	2.0089
Dimensionless Intensity B22	0.9784
Connections *per* Branch	1.4435
Trace Power-law Exponent	−1.7546

Table 3
Paired total area and circle groups with displayed counts of occurrence, aggregated from the stage 2 subsampling dataset.

Total Area m^2	1 to 4 circles	5 to 7 circles	8 to 10 circles	11 to 13 circles
0–4000	33365	4328	312	12
4000–8000	5330	17599	13729	4852
8000–20800	3	1096	9065	10309

represent lower spatial coverage.

In Fig. 8 each vertical column represents a circle count group, rows represent the four selected network parameters and within the columns the three boxplots represent the three total area groups. For all parameters, increasing circle count or total area decreases the variance of parameter values. For both count and area the increase has seemingly diminishing returns. Boxplots for all fracture network parameters show

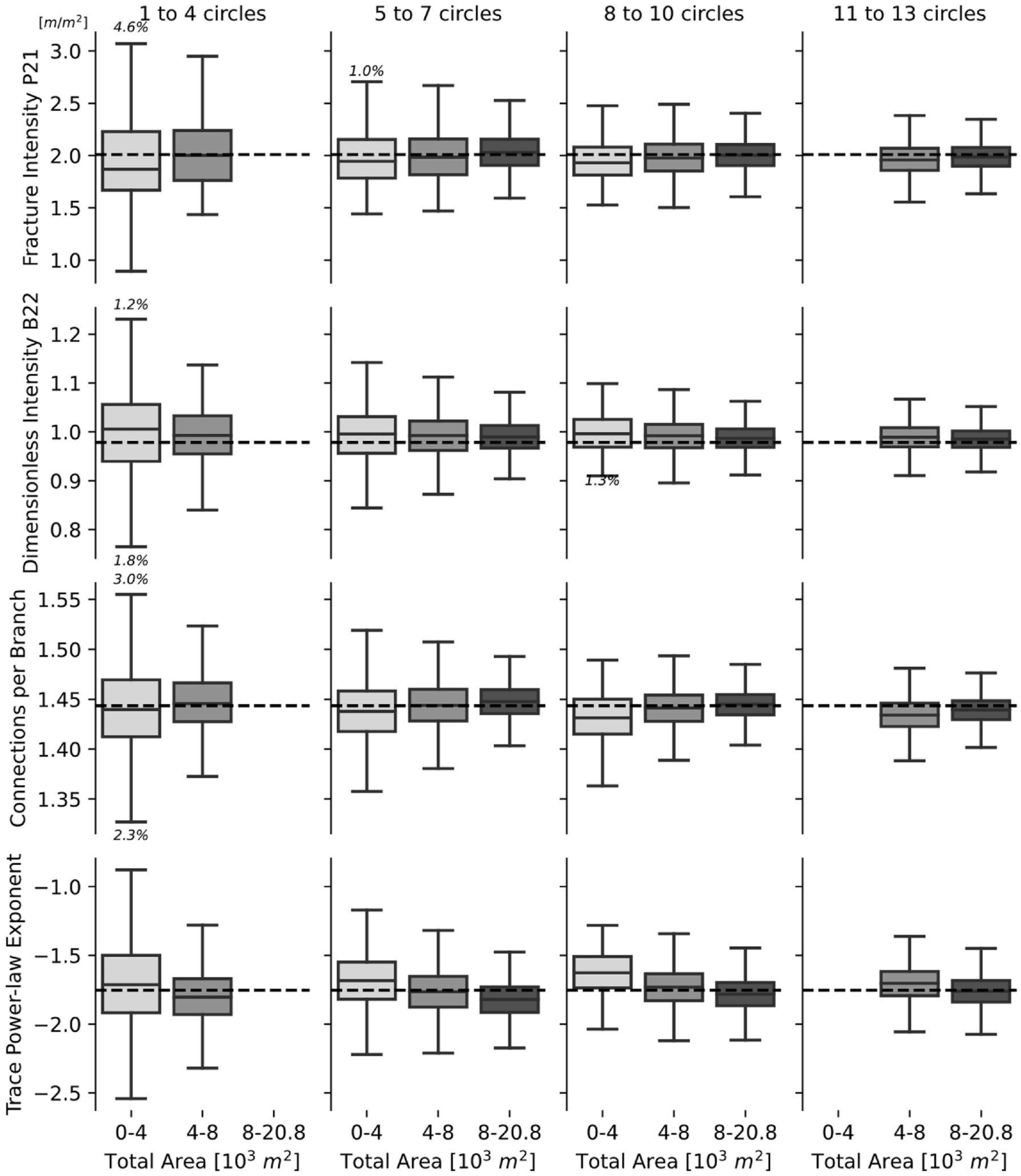

Fig. 8. We visualized the variance within each group pair with boxplots (Table 3). Each box is drawn between the first and third quartiles and whiskers are calculated with 1.5 times the inter-quartile range (Tukey, 1977). Resulting outliers are not shown to focus the plot but are instead displayed as percentages of the total quantity when the percentage is higher than 1.0%. The total area is constrained by the base circle radii and therefore very few subsample iterations had high area (8–20.8) combined with low circle count (1–4). We filter the group pairs (i.e. boxes) to only those containing at least 100 subsample iterations (=0.1% of all iterations). The dashed line for each row is at the reference value for that parameter (Table 2).

that the overall variance, indicated by boxplot whisker width, is at its highest and therefore precision at its lowest when the areal coverage is at its lowest (circle count between 1 and 4, total area between 0 and 4000 m^2; Fig. 8).

For *Connections per Branch*, *Fracture Intensity P21, Dimensionless Intensity B22* and *Trace Power-law Exponent* the median values, indicated by the continuous lines inside the boxplots, are very close to the dashed reference value lines (Table 2) in almost all boxplots (Fig. 8). However, for *Dimensionless Intensity B22* there is an apparent correlation with median value and circle count where the median value approaches the reference value as circle count and area is increased i.e. the accuracy and precision both increase. In contrast, no direct correlation between the obtained parameter accuracy and subsample circle radius or count can be detected for *Connections per Branch*, *Fracture Intensity P21* and *Trace Power-law Exponent*. For *Connections per Branch* and *Fracture Intensity P21* the median value for the parameter in all grouped subsampling boxplots (Fig. 8) is almost equivalent with its reference value (Table 2). However, the precision of the estimate based on limited sampling is lower compared to estimates based on higher sampling coverage. For *Connections per Branch* this difference in precision is the most negligible even when taking into account the limited numerical range for the parameter (0.0–2.0; Sanderson and Nixon, 2015). For *Trace Power-law Exponent* a very low precision with a total area of 0–4000 m^2 is observed. With a total sampling area of 4000–8000 m^2 (and above) the subsampled *Trace*

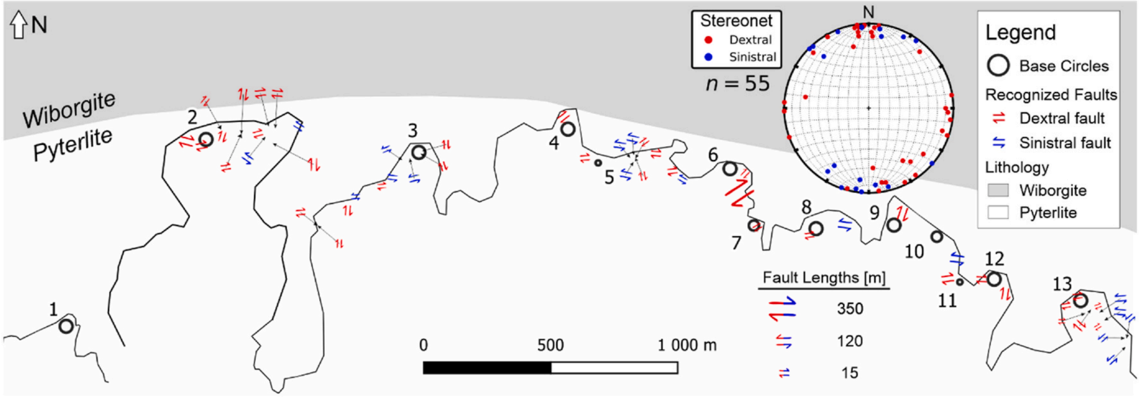

Fig. 9. The mapped faults within the Getaberget outcrops. The displayed faults are all subvertical and above 15 *m* in length. The fault symbol size correlates with the interpreted fault length that was estimated in the field or remotely from the orthophotographs with the maximum length being 350 *m* and minimum 15 *m*.

Power-law Exponent values are more precise and have a median value almost equivalent to the reference value (Fig. 8; Table 2).

4.3. Characterization of faults and fault zones

Based on field mapping, we found a total of 55 faults or fault zones, which display an overall two-fold orientation distribution: the dominant population of roughly E-W trending sub-vertical faults contains both sinistral and dextral faults, whereas the smaller N–S trending population contains only dextral faults (Fig. 9). Six to seven dextral faults define an approximately NE-SW trending cluster. The mapped faults appear to be scattered within the study area, and do not show any clear systematics with regard to their kinematic character and size and hence do not allow recognizing specific fracture domains in the scale of observation displayed in Fig. 9. With respect to the studied base circles, two circles are located within areas with the highest density of mapped faults (base circles 2 and 13) and six circles are at least partially transected by faults (2, 3, 7, 8, 12, 13), where one is affected by N–S faults (3), and others by approximately E-W trending faults. Seven of the circles (1, 4, 5, 6, 9, 10, 11) are not cut by any observed faults. The largest observed fault in the area (>350 *m* trace length) is not directly transecting any of the circles but occurs between base circles 6 and 7.

5. Discussion

The study site contains several recognized faults (Fig. 9) distributed along the outcrops. As the possible displacement within a fault typically correlates with its length and the damage zone correlates with its displacement, we can make a hypothesis that the interpreted fault length can correlate with the intensity of fracturing near the fault (Choi et al., 2016). This seems to be the case for base circle 6 which has a high *Fracture Intensity P21* value, and which is located north of the largest found fault in the area. However base circle 7 has seemingly equivalent intensity while located near the same fault possibly due to an asymmetric damage zone (Choi et al., 2016; Jokiniemi, 2021). No large fault is recognized near base circle 3 even though it has the highest *Fracture Intensity P21* value, and for this reason, the high fracture frequency likely relates to the presence of the two >15 m long N–S faults and associated fractures, with possible contribution from a larger fault that could be interpreted to be located within a topographic depression to the east of base circle 3 (Fig. 9). The proximity of multiple faults has only a very slight effect on the intensity for base circles 2 and 13, both of which are near multiple mapped faults. There is a lack of mapped faults near base circle 1 but there is no apparent correlation with the lack of faults and network parameter values (e.g. *Fracture Intensity P21* or *Connections per Branch*) when compared to base circles 11 and 12 which are near multiple faults (Figs. 7 and 9). In addition to the faults, another source of potential fracture pattern perturbation is the lithological contact between Wiborgite and Pyterlite (Fig. 9). In Fig. 6 we see an orientation maximum, WNW–ESE, of fracture traces which aligns with the contact for most of the outcrops (Fig. 9). The nature of the contact between Wiborgite and Pyterlite is unknown but as it only divides rapakivi granite variants, we expect a gradual lithological contact rather than a discrete discontinuity between the rock volumes. However, without additional data and field reconnaissance the true nature of the contact and possible effect on fracture patterns remains unknown. Similarly, the effect of faults on fracture pattern perturbation remains elusive without more efforts to document the perturbations near faults. Structural interpretations, representing a qualitative approach, could be made to attempt to characterize the brittle structural framework at the outcrops with e.g. fracture domain delineation (e.g. Follin et al., 2014; Skyttä et al., 2021). However, mapping of all possible structures (e.g. faults and lithological contacts) that affect the fracture network characteristics at the Getaberget outcrops with sufficient accuracy would be an enormous job, requiring methods to also recognize structures underwater (e.g. bathymetry) and within depressions (e.g. geophysics). Differences in trace length distributions between the base circles could be explained by the presence of faults, or by other sources of uncertainty, such as: insufficient data, censoring or spatial clustering (Nicol et al., 1996). The effect of each individual potential source of uncertainty is probably significant, and it is difficult to discriminate between the sources and estimate the magnitude of the effect of each source. As an initial solution to characterizing the numerous objective uncertainties that affect the fracture pattern at the outcrops, our subsampling method offers a numerical method to determine variations of the parameters and to determine overall characteristics more robustly than with a single-pass network characterization without any extra resources required for data collection.

Our study offers recommendations on optimizing the surveying of topological and geometric fracture network characteristics at outcrops. Based on subsampling results for *Connections per Branch* and *Fracture Intensity P21* (Fig. 8) we observe that even with very low sampling coverage (0–4000 m^2) the median parameter value for these parameters is equivalent to the value of significantly larger sampling coverage datasets (total base circle area, 20800 m^2). However, *Dimensionless Intensity B22* tends to display a higher median value compared to the reference value (Table 2; Fig. 8). Whether this represents the inadequate area and/or circle count of the subsamples or an anomaly in the full characterization reference value remains uncertain. The reference value (Table 2) might be skewed by outliers and the median value from Fig. 8 might represent a more accurate estimate of the parameter for the site. As observed in our results, smaller total area corresponds to a higher number of censored fractures which leads to a lower accuracy in power-law exponent determination (Pickering et al., 1995; Zeeb et al., 2013).

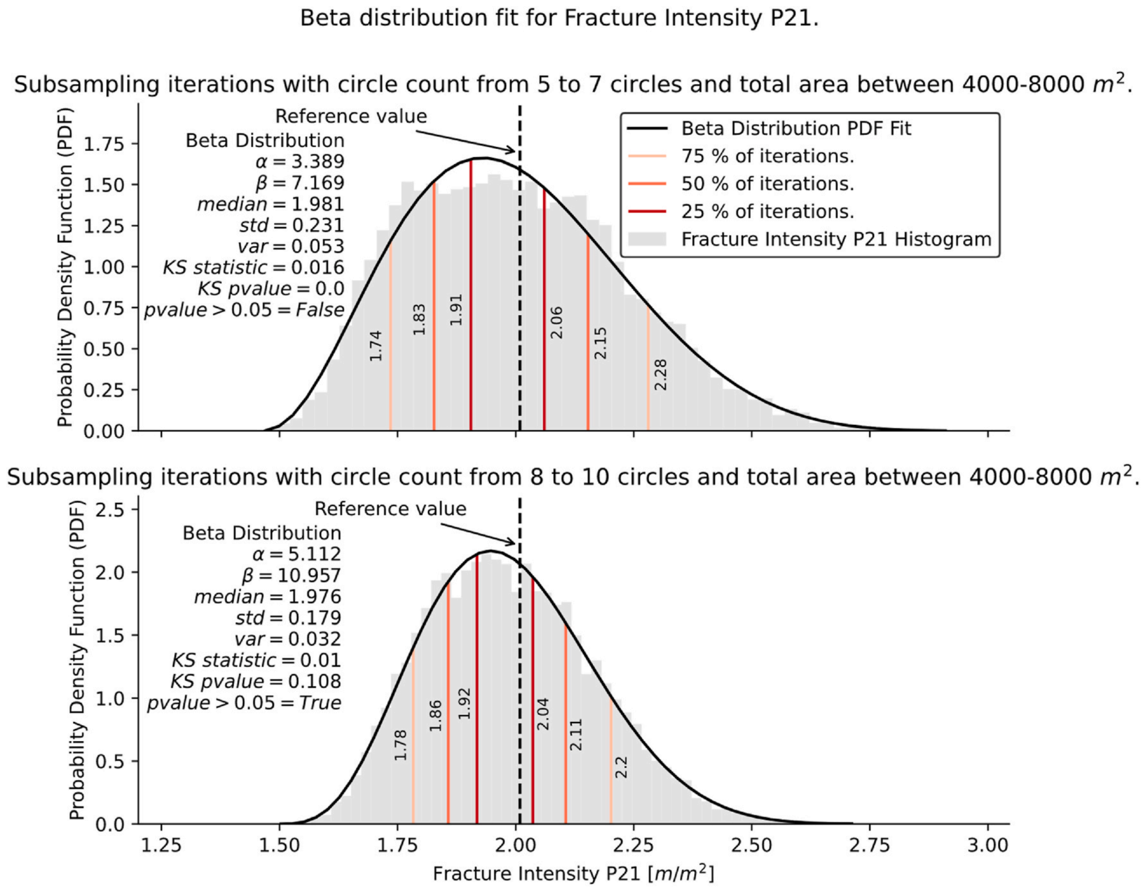

Fig. 10. Beta distribution fit for Fracture Intensity P21 values from group pairs (5–7, 4000–8000 m^2) and (8–10, 4000–8000 m^2). Reference value (Table 2) for Fracture Intensity P21 is plotted as the vertical black dashed line. The probability thresholds for values have been plotted as the red to light orange vertical lines (E.g. 25% of iterations resulted in values within the red lines). All relevant statistics and beta distribution parameters are displayed. The null hypothesis is not rejected if pvalue is greater than 0.05 (pvalue >0.05). Abbreviations: std = standard deviation, var = variance, KS statistic = Kolmogorov-Smirnov distance (*D*) and KS pvalue = Kolmogorov-Smirnov *p* value. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Based on results for *Trace Power-law Exponent*, we interpret that a total area of 8000 m^2 is sufficient for defining *Trace Power-law Exponent* accurately and with sufficient precision at the Getaberget outcrops. We make this interpretation based on firstly the visual interpretation of diminishing returns for precision when increasing the total area and secondly on the lower accuracy for exponent determination with an area between 0 and 4000. The diminishing returns is observed in the columns with 11–13 circles and 8 to 10 circles: when total area is increased from 4000 to 8000 m^2 to 8000–20800 m^2, the increase in precision is low to non-existent. Whereas when the total area is increased from 0 to 4000 m^2 to 4000–8000 m^2 the increase in precision is similarly low but with an increase in the accuracy of exponent determination. To meet this criterion of 8000 m^2 total area with 8 identical target circles, the radius of each would be approximately 18 m. We offer this recommendation as a conservative first-pass assessment based strictly on our numerical results and e.g. the use of multiple scales of observation for defining *Trace Power-law Exponent* could further increase the accuracy in power-law exponent determination (Odling, 1997; Bonnet et al., 2001; Dichiarante et al., 2020). A more rigorous geological approach could be used to point out base circles that due to lithological or structural differences could be filtered out when characterizing the fracture network, while keeping in mind that filtering adds another layer of qualitative assessment into the results. We do not perform a statistical goodness-of-fit test for the determined power-law exponents of each individual subsample and e.g. a one-sample Kolmogorov-Smirnov test could be conducted to determine statistical significance of the exponent and those with low significance could be filtered out. In addition, the other potential statistical models for fracture trace length distributions (lognormal and exponential) could be explored using the presented subsampling as notably the lognormal fit more accurately describes the length distribution (Table 1; Fig. 5). The presented Stage 1 dataset could be examined by pinpointing within each base circle the subsamples in which the goodness-of-fit of the power-law (or lognormal/exponential) fit is best to potentially reveal information about the correlation between local fracture patterns and fractal-characteristics of length distributions. The trend observed in Fig. 7 for *Trace Power-law Exponent* must also be accounted as a probable cause of added imprecision in the presented subsampling results. If all mentioned caveats related to the introduced method are considered, it is likely that an accurate assessment of *Trace Power-law Exponent* could be made from a smaller area than the recommended 8000 m^2. But, we introduce the area recommendation of 8000 m^2 with 8 or more circular sampling circles as a preliminary minimum to determine fracture trace power-law exponents accurately and precisely in comparable site studies.

We further characterize two of the group pairs (each parameter box in Fig. 8 represents a group pair) by fitting a statistical distribution to values of two parameters, *Fracture Intensity P21* and *Trace Power-law Exponent*, of both group pairs to examine the precision of their subsampling results. We picked two group pairs with circle counts of 5–7 and 8–10, both with the same total area of 4000–8000 m^2. These represent group pairs with high subsampling iteration counts (Table 3) with total area of all iterations always less than half of that of the base circles (20800 m^2) i.e. the total amount of possible subsample arrangements is very high within both. With limited testing we found that

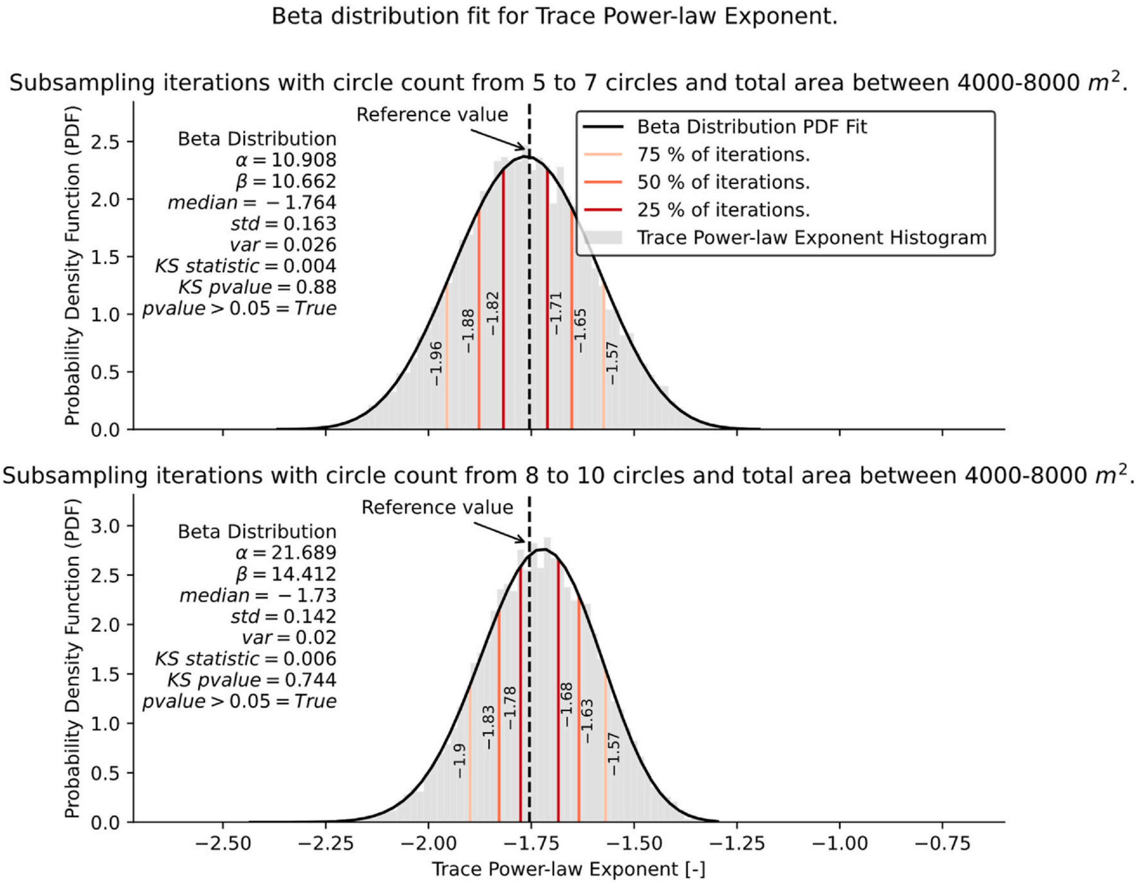

Fig. 11. Beta distribution fit for Trace Power-law Exponent values from group pairs (5–7, 4000–8000 m^2) and (8–10, 4000–8000 m^2). See Fig. 10 caption for symbol and abbreviation explanations.

Table 4
Example interpretations made from the fitted beta distributions. Certainty is the probability that we get a value in the value range in the "Value Range" column.

Certainty (%)	Parameter	Value Range	Total Area (m^2)	Circle Count
75	Fracture Intensity P21	1.78–2.20	4000–8000	8–10
75	Trace Power-law Exponent	−1.90–−1.57	4000–8000	8–10

for both group pairs and for both parameters a beta distribution (McDonald and Xu, 1995) fits reasonably well to the distribution of parameter values. We further tested the fit with a one-sample Kolmogorov-Smirnov test (Massey, 1951). Fitting and testing the distribution has been performed with scipy Python package (Virtanen et al., 2020).

Beta distribution fits for *Fracture Intensity P21* and *Trace Power-law Exponent* are presented in Figs. 10 and 11. The beta distribution fits well for *Connections per Branch* parameter (low statistic D, high p-value) but worse for *Fracture Intensity P21* and the resultant p-value for the 5–7, 4000–8000 m^2 pair fit for *Fracture Intensity P21* values is below 0.05 (Fig. 10). We therefore reject the null hypothesis that the distribution is a beta distribution and do not make any interpretations for the results of that group pair. Other group pairs pass the test and are accepted for further interpretation. From the group pairs that pass the Kolmogorov-Smirnov test we can make direct interpretations of sampling uncertainty. Example interpretations made from the fitted beta distributions (Figs. 10 and 11) are presented in Table 4.

As we demonstrated (Figs. 10 and 11), our subsampling method can be used to get the required data for making determinations of which statistical distribution the sampling results of outcrop fracture network characterization most closely follow. If the distributions of characterization parameters are accurately determined, the distribution can be used by DFN-modelling (Mäkel, 2007), especially when the modelling is based on Monte Carlo simulation which requires input variable distribution estimates (Raychaudhuri, 2008; Xu and Dowd, 2010; O'Malley et al., 2018). Furthermore, the use of subsampling in defining the characterization result is replicable: the boxplot medians in Fig. 8 can be used instead of just one characterization of the whole area to get a more statistically sound estimate of a parameter value. Alternatively, the statistical distribution, as demonstrated in Figs. 10 and 11, can be determined and exploited to get another estimate of the expected value. However, as our method is strictly two-dimensional and implemented here only on horizontal outcrop surfaces, we will not adequately sample fractures with subhorizontal dips while subvertical fractures will be oversampled (Terzaghi, 1965). This bias must be kept in mind when making interpretations beyond two dimensions i.e., when using the results in DFN-modelling.

Extrapolation of our results beyond the scale of observation used in this study (orthophotograph ground sampling distance of 0.55 $cm/pixel$) and to other lithological domains with different fracture and fault patterns requires further site-specific studies. Such studies would essentially replicate the workflow of our study while considering the site-specific geological environment. We do not constrain how the individual circle sizes vary within each area group (Table 3; Fig. 8) and all interpretations should therefore be made conservatively. However, the results of the identified parameter variance (Fig. 8) can be used in further studies at Getaberget and possibly, with caution, in comparable lithological settings elsewhere. The results are of interest when planning outcrop fracture network characterization in an initially unmapped area. Both the target area count and total area can be minimized, where the minimum is constrained by the level of acceptable precision in the characterization. We recommend comparing initial data or estimates of e.g. the fracture trace lengths in a new target area to the presented values of minimum, maximum and mean trace length values (Table 2) to assess the applicability of our recommendations. We can further differentiate between the parameters by overall variance. *Connections per Branch* has markedly low variance in Fig. 8 and, consequently, if only it is of interest in an outcrop investigation, determining the value precisely and accurately should require a much smaller total sampling area compared to e.g. *Fracture Intensity P21* even when taking into account the lower range of possible values for *Connections per Branch* (0.0–2.0; Sanderson and Nixon, 2015).

6. Conclusions

- We introduce fractopo, an open-source Python package developed for two-dimensional fracture and lineament trace map analysis, and fractopo-subsampling, Python scripts and notebooks made for the subsampling of trace networks which use and extend fractopo functionality.
- We introduce novel results of areal sampling and uncertainty analysis related to the characterization of the Getaberget outcrop fracture network by displaying the relationships between sampling coverage and characterization results.
- We recommend the use of subsampling to minimize the uncertainty related to outcrop fracture network characterization and for that purpose introduce a repeatable method, implemented in Python.
- Our results allow preliminary optimization of areal fracture network sampling coverage in crystalline rocks. E.g. when sampling outcrop fracture networks with trace lengths ranging approximately from 1 cm to 35 m from drone orthomosaics surveyed with a ground sampling distance of 0.55 $cm/pixel$ we recommend a total sampling area of 8000 m^2 with 8 or more circular sampling circles with radii of circa 18 m for accurate and precise determination of fracture trace power-law exponent.

Computer code availability

Source codes of both fractopo and fractopo-subsampling are available on GitHub in open-source repositories and fractopo is in addition available on *Python Packaging Index* (PyPI). Both are licensed with the MIT license. Code development has been done by Nikolas Ovaskainen (nikolas.ovaskainen@gtk.fi) and started in 2020 for fractopo and 2021 for fractopo-subsampling.

Installation is currently supported on Linux-based operating systems and requires Python 3.8. See fractopo-subsampling GitHub page for step-by-step guide on how to exactly replicate the subsampling environment and the results of our paper. In the future the methods implemented in fractopo-subsampling will be added to the fractopo repository and further development related to subsampling and network characterization will be implemented there. We plan to support Windows installation with the conda package manager.

- fractopo version 0.0.2 (https://github.com/nialov/fractopo/tree/0.0.2)
- fractopo-subsampling (https://github.com/nialov/fractopo-subsampling)

Data availability

All data used in this paper are freely available and permissibly licensed:

- Fracture traces and base circles (https://doi.org/10.34740/kaggle/dsv/2101134) which can be opened in any GIS-software for inspection of digitized fracture traces.
- Subsampled dataset (https://doi.org/10.34740/kaggle/dsv/2926127) as csv-files which can be used in any statistical analysis software to verify statistical subsampling results.
- Drone orthomosaics (https://doi.org/10.5281/zenodo.4719627) which can be opened in any GIS-software for inspection and cross-referencing with the digitized fracture traces.
- Recognized fault locations and attributes as a csv-file with coordinates in ETRS-TM35FIN, EPSG:3067 (https://doi.org/10.5281/zenodo.5746625).

Table 5
Uniquely ID-labeled base circles with radii and the filenames that are associated to the raster filenames.

ID	Radius [m]	Name
1	25	Getaberget_20m_4_3_area
2	25	Getaberget_20m_9_2_area
3	25	Getaberget_20m_8_3_area
4	25	Getaberget_20m_7_1_area
5	10	Getaberget_20m_7_2_area
6	25	Getaberget_20m_5_1_area
7	20	Getaberget_20m_2_1_area
8	25	Getaberget_20m_2_2_area
9	25	Getaberget_20m_1_1_area
10	20	Getaberget_20m_1_2_area
11	10	Getaberget_20m_1_3_area
12	25	Getaberget_20m_1_4_area
13	25	Havsvidden_20m_1_area

To identify the drone orthomosaic rasters by their ID the base circles are catalogued by their raster-linked filenames in Table 5. See Fig. 1 for base circle locations.

Authorship statement

Nikolas Ovaskainen: Conceptualization, Methodology, Software, Investigation, Writing - Original Draft **Nicklas Nordbäck**: Methodology, Investigation, Writing - Review & Editing **Pietari Skyttä**: Methodology, Writing - Review & Editing **Jon Engström**: Writing - Review & Editing, Investigation.

Nikolas Ovaskainen is responsible for all code, most of the method development and is the main writer contributing to drafting of all sections. Nicklas Nordbäck contributes in method development and writing in the methods section, in interpretation of results, data acquisition and drafting in all other sections with comments and suggestions. Pietari Skyttä contributes primarily with drafting, commenting and revision suggestions in the introduction and discussion sections and the abstract and secondarily with result interpretation. Jon Engström contributes with draft commenting and revision suggestions in all sections.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the Finnish Research Programme on Nuclear Waste Management (2019–2022) and Geological Survey of Finland for funding the KYT KARIKKO -project, Ismo Aaltonen for digitizing two target areas, the constructive reviews by David Sanderson and an anonymous reviewer and the editorial handling by Virginia Toy.

References

Agisoft, 2020. Agisoft Metashape user manual. Agisoft Metashape 160.

Alstott, J., Bullmore, E., Plenz, D., 2014. Powerlaw: a python package for analysis of heavy-tailed distributions. PLoS One 9, 1–18. https://doi.org/10.1371/journal.pone.0085777.

Alvarez, L.L., Guimarães, L.J., do, N., Gomes, I.F., Beserra, L., Pereira, L.C., Miranda, T.S. de, Maciel, B., Barbosa, J.A., 2021. Impact of fracture topology on the fluid flow behavior of naturally fractured reservoirs. Energies 14 (5488). https://doi.org/10.3390/en14175488.

Andrews, B., Roberts, J., Shipton, Z., Bigi, S., Chiara Tartarello, M., Johnson, G., 2019. How do we see fractures? Quantifying subjective bias in fracture data collection. Solid Earth 10, 487–516. https://doi.org/10.5194/se-10-487-2019.

Bemis, S.P., Micklethwaite, S., Turner, D., James, M.R., Akciz, S., Thiele, S.T., Bangash, H.A., 2014. Ground-based and UAV-Based photogrammetry: a multi-scale, high-resolution mapping tool for structural geology and paleoseismology. J. Struct. Geol. 69, 163–178. https://doi.org/10.1016/j.jsg.2014.10.007.

Bertrand, L., Géraud, Y., Le Garzic, E., Place, J., Diraison, M., Walter, B., Haffen, S., 2015. A multiscale analysis of a fracture pattern in granite: a case study ofthe Tamariu granite, Catalunya, Spain. J. Struct. Geol. 78, 52–66. https://doi.org/10.1016/j.jsg.2015.05.013.

Bieniawski, Z.T., Denkhaus, H.G., Vogler, U.W., 1969. Failure of fractured rock. Int. J. Rock Mech. Min. Sci. Geomech. Abstracts 6, 323–341. https://doi.org/10.1016/0148-9062(69)90009-6.

Bisdom, K., Nick, H.M., Bertotti, G., 2017. An integrated workflow for stress and flow modelling using outcrop-derived discrete fracture networks. Comput. Geosci. 103, 21–35. https://doi.org/10.1016/j.cageo.2017.02.019.

Bond, C.E., Gibbs, A.D., Shipton, Z.K., Jones, S., 2007. What do you think this is? "Conceptual uncertainty" In geoscience interpretation. GSA Today (Geol. Soc. Am.) 17, 4–10. https://doi.org/10.1130/GSAT01711A.1.

Bonnet, E., Bour, O., Odling, N.E., Davy, P., Main, I., Cowie, P., Berkowitz, B., 2001. Scaling of fracture systems in geological media. Rev. Geophys. 39, 347–383.

Choi, J.H., Edwards, P., Ko, K., Kim, Y.S., 2016. Definition and classification of fault damage zones: a review and a new methodological approach. Earth Sci. Rev. 152, 70–87. https://doi.org/10.1016/j.earscirev.2015.11.006.

Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Rev. 51, 661–703. https://doi.org/10.1137/070710111.

Dershowitz, W.S., Herda, H.H., 1992. Interpretation of fracture spacing and intensity. In: The 33rd u.S. Symposium on Rock Mechanics (USRMS). American Rock Mechanics Association, Santa Fe. New Mexico, 10.

Dichiarante, A., McCaffrey, K., Holdsworth, R., Bjornarå, T., Dempsey, E., 2020. Fracture attribute scaling and connectivity in the Devonian Orcadian Basin with implications for geologically equivalent sub-surface fractured reservoirs. Solid Earth Discussions 1–48. https://doi.org/10.5194/se-2020-15.

Follin, S., Hartley, L., Rhén, I., Jackson, P., Joyce, S., Roberts, D., Swift, B., 2014. A methodology to constrain the parameters of a hydrogeological discrete fracture network model for sparsely fractured crystalline rock, exemplified by data from the proposed high-level nuclear waste repository site at Forsmark, Sweden. Hydrogeol. J. 22, 313–331. https://doi.org/10.1007/s10040-013-1080-2.

GTK, 2017. Bedrock of Finland 1 (200), 000.

Haapala, I., Rämö, O.T., 1992. Tectonic setting and origin of the Proterozoic rapakivi granites of southeastern Fennoscandia. Trans. R. Soc. Edinb. Earth Sci. 83, 165–171. https://doi.org/10.1017/S0263593300007859.

Hansberry, R.L., King, R.C., Holford, S.P., Hand, M., Debenham, N., 2021. How wide is a fault damage zone? Using network topology to examine how fault-damage zones overprint regional fracture networks. J. Struct. Geol. 146 https://doi.org/10.1016/j.jsg.2021.104327.

Hardebol, N.J., Bertotti, G., 2013. DigiFract: a software and data model implementation for flexible acquisition and processing of fracture data from outcrops. Comput. Geosci. 54, 326–336. https://doi.org/10.1016/j.cageo.2012.10.021.

Healy, D., Rizzo, R.E., Cornwell, D.G., Farrell, N.J.C., Watkins, H., Timms, N.E., Gomez-Rivas, E., Smith, M., 2017. FracPaQ: a MATLAB™ toolbox for the quantification of fracture patterns. J. Struct. Geol. 95, 1–16. https://doi.org/10.1016/j.jsg.2016.12.003.

James, M.R., 2017. SfM-MVS PhotoScan image processing exercise. IAVCEI, 2017.

Jordahl, K., Bossche, J.V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A.G., Farmer, C., Hjelle, G.A., Snow, A.D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., Maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L.J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, Leblanc, F., 2020. Geopandas/geopandas: v0.8.1. Zenodo. https://doi.org/10.5281/zenodo.3946761.

Jokiniemi, J., 2021. 3D-modelling of fault-induced small-scale secondary fracturing in crystalline rocks (Master's Thesis). University of Turku, Turku. https://urn.fi/URN:NBN:fi-fe2021080642375.

Kim, Y.S., Peacock, D.C.P., Sanderson, D.J., 2004. Fault damage zones. J. Struct. Geol. 26, 503–517. https://doi.org/10.1016/j.jsg.2003.08.002.

Laitakari, L., Rämö, T., Suominen, V., Niin, M., Stepanov, K., Amantov, A., 1996. Subjotnian: rapakivi granites and related rocks in the surroundings of the Gulf of Finland. Spec. Pap. Geol. Surv. Finland 21, 59–97.

Loza Espejel, R., Alves, T.M., Blenkinsop, T.G., 2020. Multi-scale fracture network characterisation on carbonate platforms. J. Struct. Geol. 140, 104160 https://doi.org/10.1016/j.jsg.2020.104160.

Luosto, U., Tiira, T., Korhonen, H., Azbel, I., Burmin, V., Buyanov, A., Kosminskaya, I., Ionkis, V., Sharov, N., 1990. Crust and upper mantle structure along the DSS Baltic profile in SE Finland. Geophys. J. Int. 101, 89–110. https://doi.org/10.1111/j.1365-246X.1990.tb00760.x.

Manzocchi, T., 2002. The connectivity of two-dimensional networks of spatially correlated fractures. Water Resour. Res. 38, 1–20. https://doi.org/10.1029/2000wr000180, 1–1.

Massey, F.J., 1951. The Kolmogorov-smirnov test for goodness of fit. J. Am. Stat. Assoc. 46, 68–78. https://doi.org/10.1080/01621459.1951.10500769.

Mauldon, M., 1998. Estimating mean fracture trace length and density from observations in convex windows. Rock Mech. Rock Eng. 31, 201–216. https://doi.org/10.1007/s006030050021.

Mauldon, M., Dunne, W.M., Rohrbaugh, M.B., 2001. Circular scanlines and circular windows: new tools for characterizing the geometry of fracture traces. J. Struct. Geol. 23, 247–258. https://doi.org/10.1016/S0191-8141(00)00094-8.

Mäkel, G.H., 2007. The modelling of fractured reservoirs: constraints and potential for fracture network geometry and hydraulics analysis. Geol. Soc. Spec. Publ. 292, 375–403. https://doi.org/10.1144/SP292.21.

McDonald, J.B., Xu, Y.J., 1995. A generalization of the beta distribution with applications. J. Econom. 66, 133–152. https://doi.org/10.1016/0304-4076(94)01612-4.

Mooney, C.Z., 1997. Monte Carlo Simulation. Sage Publications, Thousand Oaks, Calif.

Nelson, R.A., 1985. Geologic Analysis of Naturally Fractured Reservoirs, first ed. Gulf Publishing Co., Houston.

Nickelsen, R.P., Hough, V.N.D., 1967. Jointing in the appalachian plateau of Pennsylvania. Geol. Soc. Am. Bull. 78, 609–630.

Nicol, A., Walsh, J.J., Watterson, J., Gillespie, P.A., 1996. Fault size distributions — are they really power-law? J. Struct. Geol. 18, 191–197. https://doi.org/10.1016/S0191-8141(96)80044-7.

Nironen, M., 1997. The Svecofennian Orogen: a tectonic model. Precambrian Res. 86, 21–44. https://doi.org/10.1016/S0301-9268(97)00039-9.

Nixon, C.W., Nærland, K., Rotevatn, A., Dimmen, V., Sanderson, D.J., Kristensen, T.B., 2020. Connectivity and network development of carbonate-hosted fault damage zones from western Malta. J. Struct. Geol. 141 https://doi.org/10.1016/j.jsg.2020.104212, 104212.

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. https://doi.org/10.1130/GES01595.1.

O'Malley, D., Karra, S., Hyman, J.D., Viswanathan, H.S., Srinivasan, G., 2018. Efficient Monte Carlo with graph-based subsurface flow and transport models. Water Resour. Res. 54, 3758–3766. https://doi.org/10.1029/2017WR022073.

Odling, N.E., 1997. Scaling and connectivity of joint systems in sandstones from western Norway. J. Struct. Geol. 19, 1257–1271. https://doi.org/10.1016/S0191-8141(97)00041-2.

Ortega, O.J., Marrett, R.A., Laubach, S.E., 2006. A scale-independent approach to fracture intensity and average spacing measurement. AAPG (Am. Assoc. Pet. Geol.) Bull. 90, 193–208. https://doi.org/10.1306/08250505059.

Ovaskainen, N., 2020. Scalability of lineament and fracture networks within the crystalline Wiborg Rapakivi Batholith, SE Finland. University of Turku. Master's Thesis.

Pahl, P.J., 1981. Estimating the mean length of discontinuity traces. Int. J. Rock Mech. Min. Sci. Geomech. Abstracts 18, 221–228. https://doi.org/10.1016/0148-9062(81)90976-1.

Palamakumbura, R., Krabbendam, M., Whitbread, K., Arnhardt, C., 2020. Data acquisition by digitizing 2-D fracture networks and topographic lineaments in geographic information systems: further development and applications. Solid Earth 11, 1731–1746. https://doi.org/10.5194/se-11-1731-2020.

Peacock, D.C.P., Sanderson, D.J., Bastesen, E., Rotevatn, A., Storstein, T.H., 2019. Causes of bias and uncertainty in fracture network analysis. Norw. J. Geol. 99, 1–16. https://doi.org/10.17850/njg99-1-06.

Pickering, G., Bull, J.M., Sanderson, D.J., 1995. Sampling power-law distributions. Tectonophysics 248, 1–20. https://doi.org/10.1016/0040-1951(95)00030-Q.

Prabhakaran, R., Bruna, P.-O., Bertotti, G., Smeulders, D., 2019. An automated fracture trace detection technique using the complex shearlet transform. Solid Earth Discussions 1–40. https://doi.org/10.5194/se-2019-104.

Priest, S.D., 1993. Discontinuity Analysis for Rock Engineering. Dordrecht, Springer Netherlands. https://doi.org/10.1007/978-94-011-1498-1.

Priest, S.D., Hudson, J.A., 1981. Estimation of discontinuity spacing and trace length using scanline surveys. Int. J. Rock Mech. Min. Sci. Geomech. Abstracts 18, 183–197. https://doi.org/10.1016/0148-9062(81)90973-6.

Procter, A., Sanderson, D.J., 2018. Spatial and layer-controlled variability in fracture networks. J. Struct. Geol. 108, 52–65. https://doi.org/10.1016/J.JSG.2017.07.008.

Puress, K., Narasimhan, T., 1982. A Practical Method for Modeling Fluid and Heat Flow In Fractured Porous. Sixth SPE Symposium on Reservoir Simulation. Lawrence Berkeley National Laboratory, New Orleans, 42.

QGIS Development Team, 2020. QGIS geographic information system. Open Source Geospatial Found.

Rawnsley, K.D., Rives, T., Petti, J.-P., Hencher, S.R., Lumsden, A.C., 1992. Joint development in perturbed stress fields near faults. J. Struct. Geol. 14, 939–951. https://doi.org/10.1016/0191-8141(92)90025-R.

Raychaudhuri, S., 2008. Introduction to Monte Carlo Simulation. In: 2008 Winter Simulation Conference. IEEE.

Rämö, O.T., Haapala, I., 2005. Chapter 12 rapakivi granites. Dev. Precambrian Geol. 14, 533–562. https://doi.org/10.1016/S0166-2635(05)80013-1.

Rohrbaugh, M.B., Dunne, W.M., Mauldon, M., 2002. Estimating fracture trace intensity, density, and mean length using circular scan lines and windows. AAPG (Am. Assoc. Pet. Geol.) Bull. 86, 2089–2104.

Sanderson, D.J., Nixon, C.W., 2018. Topology, connectivity and percolation in fracture networks. J. Struct. Geol. 115, 167–177. https://doi.org/10.1016/j.jsg.2018.07.011.

Sanderson, D.J., Nixon, C.W., 2015. The use of topology in fracture network characterization. J. Struct. Geol. 72, 55–66. https://doi.org/10.1016/j.jsg.2015.01.005.

Sanderson, D.J., Peacock, D.C.P., 2020. Making rose diagrams fit-for-purpose. Earth-Sci. Rev. 201 https://doi.org/10.1016/j.earscirev.2019.103055.

Skyttä, P., Ovaskainen, N., Nordbäck, N., Engström, J., Mattila, J., 2021. Fault-induced mechanical anisotropy and its effects on fracture patterns in crystalline rocks. J. Struct. Geol., 104304 https://doi.org/10.1016/j.jsg.2021.104304.

Tannert, C., Elvers, H., Jandrig, B., 2007. The ethics of uncertainty. EMBO Rep. 8, 892–896. https://doi.org/10.1038/sj.embor.7401072.

Terzaghi, R.D., 1965. Sources of error in joint surveys. Geotechnique 15, 287–304. https://doi.org/10.1680/geot.1965.15.3.287.

Thiele, S.T., Grose, L., Samsu, A., Micklethwaite, S., Vollgger, S.A., Cruden, A.R., 2017. Rapid, semi-automatic fracture and contact mapping for point clouds, images and geophysical data. Solid Earth 8, 1241–1253. https://doi.org/10.5194/se-8-1241-2017.

Tukey, J.W., 1977. Exploratory data analysis.

Vasuki, Y., Holden, E.J., Kovesi, P., Micklethwaite, S., 2014. Semi-automatic mapping of geological Structures using UAV-based photogrammetric data: an image analysis approach. Comput. Geosci. 69, 22–32. https://doi.org/10.1016/j.cageo.2014.04.012.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S.J. van der, Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., Mulbregt, P. van, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Waskom, M., 2021. Seaborn: statistical data visualization. Zenodo. https://doi.org/10.21105/joss.03021.

Watkins, H., Bond, C.E., Healy, D., Butler, R.W.H., 2015. Appraisal of fracture sampling methods and a new workflow tocharacterise heterogeneous fracture networks at outcrop. J. Struct. Geol. 72, 67–82. https://doi.org/10.1016/j.jsg.2015.02.001.

Weiss, M., 2008. Techniques for estimating fracture size: a comparison of methods. Int. J. Rock Mech. Min. Sci. 45, 460–466. https://doi.org/10.1016/j.ijrmms.2007.07.010.

Xu, C., Dowd, P., 2010. A new computer code for discrete fracture network modelling. Comput. Geosci. 36, 292–301. https://doi.org/10.1016/j.cageo.2009.05.012.

Zeeb, C., Gomez-Rivas, E., Bons, P.D., Blum, P., 2013. Evaluation of Sampling methods for fracture network characterization using Outcrops. AAPG (Am. Assoc. Pet. Geol.) Bull. 97, 1545–1566. https://doi.org/10.1306/02131312042.