SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Earth and Environmental Sciences

PhD program in Chemical, Geological and Environmental Sciences Cycle XXXVIII

Curriculum of Geology

Characterization and Modelling of Fracture Networks with Applications to the Circulation and Storage of Geofluids

Casiraghi Stefano

Registration number: 813592

Tutor: Agliardi Federico

Supervisor: Andrea Bistacchi

Co-tutor: Silvia Mittempergher

Coordinatore / Coordinator: Marco Giovanni Malusà

ANNO ACCADEMICO / ACADEMIC YEAR 2024/2025

Abstract

Fracture networks play a key role in controlling fluid flow, rock stability, and numerous geological processes, yet their characterization in the subsurface is limited by the resolution of borehole and geophysical datasets. In contrast, extensive datasets can be acquired from outcrops by combining direct field observations with remote sensing techniques, such as digital outcrop models (DOMs). These data serve as a basis for constraining geometrical models of subsurface fracture networks within the outcrop analogue approach. In this framework, we developed a comprehensive workflow that spans from outcrop selection to data collection and analysis, addressing all key parameters required for the construction of stochastic Discrete Fracture Network (DFN) models. The workflow integrates two types of digital outcrop models derived from photogrammetric acquisition and reconstruction: point cloud DOMs (PC-DOMs), which are dense 3D point sets containing XYZ coordinates and RGB values, and textured surface DOMs (TS-DOMs), which are obtained by meshing point clouds and draping high-resolution images over their surfaces. The proposed workflow addresses all the main geometrical parameters required for fracture network characterization, including orientation parameters, topological attributes, length and height distribution parameters, the height/length (H/L) ratio, and the Representative Elementary Volume/Area (REV/REA) for P21 (fracture intensity).

The complete workflow is described in Chapter 1. Orientation data are collected with a semi-automatic procedure applied to PC-DOMs of the vertical side of the outcrop to extract 2D polygonal facets. Fracture sets are defined with a clustering procedure and different orientation distributions are fitted and tested with goodness-of-fit tests.

The topological analysis includes the classification of nodes into I (isolated), Y (abutting), X (crosscutting) and B (boundary), the calculation of the connectivity index (CI), and the extraction of the backbone, the largest cluster of connected fractures in the network. A new classification scheme called “directional topology” is also proposed, associating each node with its topological type and corresponding fracture set.

H/L ratio has been calculated assuming that height and length measurements on two adjacent and perpendicular sides of the outcrop are correlated, with the longest fractures being also the tallest.

Chapter 2 is dedicated to the length distribution fitting and parameters calculation. Stochastic models requires a fully specified statistical distribution, this means that the distribution type, and the relative parameters are needed. This motivated the need to develop a new methodology. Our methodology aims specifically at treating censoring bias and obtaining an unbiased trace length statistical model. We propose to tackle the censoring bias by applying survival analysis techniques: a branch of statistics focused on modelling time-to-event data and correctly estimating model parameters with data affected by censoring. Secondly, we implemented a ranking system based on statistical distances to select the most representative parametric model.

Chapter 3 has been dedicated to the problem of Representative Elementary Area (REA) definition and calculation. Here we consider REA as a range, and not as a single threshold. The methodology is based on a series of formal statistical tests and diagnostic plots to compare the mean, variance and shape of P21 distributions collected with progressively increasing scan areas.

To test all these methodologies, we considered different natural case studies: Pontrelli abandoned quarry located in the Murge Plateau near Altamura (Puglia, Italy), in the forebulge of the Southern Apennines fold and thrust belt. This case study has been the subject of different works in scientific literature, one of the reasons is the combination between a wide and clean pavement (around 18.000 m^2) where fractures are well exposed, and the associated vertical walls. This configuration enables the characterization of both horizontal and vertical fracture traces and the integration of information obtained from the two complementary types of outcrops. Colle Salza outcrop is located in the basement of the Western Alps on paragneiss of the Monte Rosa Nappe. This outcrop represents a less ideal situation with respect to Pontrelli, but certainly a more common occurrence. The boundary geometry is highly concave thus leading to a high censoring fraction of fracture traces, making it an excellent case study to test the performance of the length distribution fitting method. Lilstock outcrop is located in the southern coast of the Bristol Channel in West Somerset (UK). This outcrop has been the subject of many study in scientific literature due to the particular cleanliness and the extremely dense fracture network, resulting from the combination of eight different fracture sets. We employed an open-source dataset available as supplementary material, to test our P21 REA method.

General introduction and thesis overview

Fracture networks play a crucial role in a wide range of geological and engineering applications. They control fluid circulation and permeability in various types of reservoirs (March et al., 2017; Wallace et al., 2021; Wang et al., 2022; Forstner et al., 2025) and influence numerous Earth surface processes (Eppes et al., 2024). Their role extends to geotechnical and engineering geology applications (Agliardi et al., 2017; Eberhardt et al., 2004; Franzosi et al., 2023a, b) such as slope stability assessment, as well as to hydrogeological problems like contaminant transport (Cherubini, 2008; Medici et al., 2024). These broad applications have justified the growing interest in accurately characterizing and modelling these geological features (Eppes et al., 2024).

When it comes to reservoir modelling and fluid flow simulations, fractures in the subsurface can only be partially characterized at the mesoscale (from meters to tens of meters) using direct techniques. Boreholes provide local, one-dimensional information within a three-dimensional volume, including fracture orientation, aperture, abundance (P10; Dershowitz and Herda, 1992), and, when properly oriented, the one-dimensional spatial arrangement of fractures. However, parameters such as length and height distributions, connectivity, and the representative elementary area or volume (REA/REV) cannot be directly measured.

Geophysical methods can offer continuous three-dimensional data, but they are subject to important limitations: (i) fractures are not always associated with sufficient contrasts in physical properties to be imaged, and (ii) the spatial resolution of such datasets is inherently restricted. For instance, in high-quality industrial 3D seismic surveys, fractures smaller than approximately 200–300 m cannot be resolved. Moreover, to be directly imaged, a fracture must exhibit a displacement that produces a detectable seismic impedance contrast across the discontinuity. Consequently, only macro-scale faults can be reliably imaged in seismic data, leading to biased estimates of volumetric fracture metrics (e.g., Laubach et al., 2019).

Given these limitations, the outcrop analogue approach is commonly adopted to provide input properties for generating stochastic Discrete Fracture Network (DFN) models. This approach assumes that detailed information acquired from well-exposed outcrops can be used as a representative analogue of the fracture network properties in the subsurface, thereby compensating for the lack of direct reservoir-scale data. The applicability of an outcrop as an analogue must, however, be carefully evaluated, and certain assumptions are often required (Forstner and Laubach, 2022). Fracture networks are therefore a complex geological object whose full characterization, connection to subsurface reservoirs and modelling require multi-disciplinary skills that go from statistics (Mauldon et al., 2001; Zeeb et al., 2013; Healy et al., 2017), geochemistry (Beaudoin et al., 2022; Berio et al., 2022; Cruset et al., 2023), and structural geology and involve the determination of several parameters (Table 1).

Parameter	Fracture network	Fracture set	DOM - Facets	DOM - Traces
Number of sets	*		*	*
Orientation		*	*	
Topology	*	*		*
Size (length/height)		*		*
H/L ratio		*		*
Density/Intensity	*	*		*
Aperture		*		
Spatial organization	*	*		*
Representative Elementary Volume, Area	*	*		*
Roughness		*	*	*
Kinematics		*		
Deformation Mechanism		*		
Filling		*		

Table 1 Summary of the parameters necessary for quantitative characterization of fracture networks. Fracture network column indicates whether it is possible to calculate the parameter on the entire fracture network. Fracture set column indicates whether it is possible to calculate the parameter on a single fracture set. DOM – Facets indicates if it is possible to calculate the parameter on a fracture plane

In this context, the main objective of this doctoral thesis is to improve the characterization of key geometrical parameters of fracture networks from outcrop analogues with statistically sound methodologies, with the aim of generating more realistic stochastic models and validating model results. The research addresses several open issues, including biases affecting data collected from outcrops, the H/L ratio problem, and the determination of the REA range for the P21 parameter. The thesis is structured into three main chapters, each organized in the form of a scientific paper: the first chapter has been accepted for publication, the second has already been published, and the third is presented as a draft manuscript.

Chapter 1

The first chapter outlines the overall methodological workflow and provides the most general framework for the study. This chapter describes a general workflow for the characterization of all the geometrical parameters needed to produce a fracture network 3D stochastic model. In the chapter introduction is presented a general overview of the open problems when comes to the fracture network characterization from outcrops. It follows a section that describes the best practice to choose an optimal outcrop for this kind of analysis and a section regarding the Digital

Outcrop Models (DOMs) reconstruction and pre-processing. These sections are of particular importance because all the subsequent analysis and methodologies, even the other two chapters, are based on DOMs. There are two main products resulting from photogrammetric processing: (i) Point cloud DOMs (PC-DOMs) that are a dense sets of points, where each point is characterized by XYZ coordinates and an RGB value, and (ii) textured surface DOMs (TS-DOMs) that are derived from PC-DOMs by generating a polygonal mesh from the point cloud and texturing images onto its surface.

Starting from Section 4, a semi-automatic methodology for collecting orientation data from PC-DOMs is presented. This approach minimizes interpreter bias by combining cluster analysis with traditional structural observations and provides a statistically grounded assessment of fracture set orientation parameters through the implementation of a goodness-of-fit test for the Fisher distribution. A calibrated semi-automatic data collection step is also introduced to increase the amount of data available for statistical analysis.

Sections 5 and 6 focus on the fracture network topological analysis. As part of this doctoral project, a new Python library dedicated to topological analysis was developed (available at https://github.com/gecos-lab/FracAbility). The analysis is based on data obtained from TS-DOMs. Compared to standard approaches that classify nodes according to the relative proportions of I, Y, and X types, we integrated the detection of B (boundary) nodes. These nodes, resulting from intersections between fractures and the interpretation boundary, must be excluded from the analysis to avoid underestimating the connectivity index. Furthermore, we implemented the extraction of the fracture network backbone, which corresponds to the largest connected component within the network, and introduced the concept of directional topology, in which each I, Y, and X node is associated with its corresponding fracture set (I nodes), or sets (Y and X nodes). This approach enables a more detailed classification and provides more targeted input parameters for stochastic modelling.

Section 7 is dedicated to the statistical fitting of fracture length and height distributions, applying the concept of survival analysis to address the problem of censoring bias. The methodology is presented here in a summarized form and will be discussed in detail in the following chapter.

Section 8 presents a methodology for defining the representative elementary area (REA) of the P21 parameter. P21 values were obtained using hexagonal grids of sampling windows with progressively increasing sizes. In contrast to other methodologies proposed in the scientific literature, we adopted a more qualitative approach based on two parameters: (i) the delta interquartile range (delta IQR), calculated as the difference between the interquartile ranges of two consecutive P21 samples; and (ii) to account for outliers data not included in the IQR, the range between the whiskers, defined as the difference between the upper and lower whisker values.

In Section 9 H/L ratio problem is addressed. This parameter is fundamental in the stochastic modelling context, because it is used to upscale parameters collected in 2D (outcrop surface) in a 3D volume. It is not possible to

deterministically associate vertical traces with horizontal traces to calculate this parameter, an assumption is always required. Our assumption is that traces mapped on the pavement (i.e. lengths) and on the wall (i.e. heights) can be associated in ordered pairs from the shortest to the longest. To test this assumption a hundred values of length and height are randomly sampled from the statistical distributions of length and height, ordered from smallest to largest and associated in pairs, and the H/L ratio is obtained with linear regression. Hundred realizations are made to account for the variability of the random sampling, and the average H/L ratio is kept as the representative value.

Chapter 2

Although included as the second chapter, this study was the first to be completed chronologically during the PhD. For this reason, it closely overlaps with Section 7 of the first chapter, where the same work is presented in a summarized form. The main focus of this chapter is the development of a new methodology to correct for censoring bias and obtain unbiased fracture length statistics.

When measuring fracture lengths from outcrops, four major sources of bias must be considered: orientation bias, truncation bias, censoring bias, and size bias. Censoring bias arises naturally due to the finite extent of the outcrop or the presence of obstacles such as vegetation or debris. While the other biases can be effectively addressed—or at least quantified—by selecting outcrops with at least two perpendicular exposed surfaces (orientation bias), defining a truncation limit (truncation bias), or adapting the sampling strategy (size bias), censoring bias cannot be simply resolved during data acquisition, as it is inherently linked to the limited extent of the outcrop itself. Previous studies have attempted to address censoring bias, for instance through the use of circular scan area or circular scan line methods (Mauldon et al., 2001).

The problem is that analysis carried out with the circular scanline method yield estimates of mean length values but are unable to provide a complete characterization of the lineament length distribution and therefore lack any real statistical significance. This is especially true considering that stochastic models require as input a fully specified statistical distribution. The methodology presented in this chapter represents a powerful alternative to circular scanline methods to treat specifically the censoring bias and obtain unbiased trace length statistical models. From a statistical point of view, this problem is analogous to the censoring bias affecting some medical, biological, and engineering datasets, and the techniques used in these disciplines to solve or limit the effects of this bias go under the names of survival analysis, life testing, or reliability analysis (Kaplan and Meier, 1958; Leung et al., 1997; Lawless, 2003; Cox, 2017; Karim and Islam, 2019). Even though in these disciplines the recorded random variables are time spans (e.g., lifetime of a patient, time-to-failure of a mechanical part, etc.), we demonstrate that this statistical technique can be also adapted to length measurements. Survival/reliability analysis techniques are adapted to correctly account for censored lengths and estimate robust trace length distributions derived from DOMs. In the

end a quantitative methodology to select the most representative estimated statistical model (i.e. parametric distribution) from a list of proposed models is proposed.

Chapter 3

This chapter is the last work developed during the PhD project, and it is a direct evolution of the REA range qualitative estimation proposed in section 8 of chapter 1. When comes to REV (Representative elementary Volume) or REA (Representative Elementary Area) calculation, it has been demonstrated that there is not a unique REV for a specific fracture network, but different REV can be identified for different parameters. The REV concept has been widely applied to different fields of applications, and is often defined as a single threshold that marks the transition between a discontinuous behavior to a continuous one, where the target parameter value becomes independent from the position at which it is calculated. In classical theoretical REV theory representation (Bear, 1975; Hudson and Harrison, 1997) another transition to an inhomogeneous behavior is expected when large scale structures are progressively included in the sampling element (1D, 2D or 3D). This upper limit has received much less attention in scientific literature with respect to the lower limit that has been thoroughly investigated for different parameters in different fields of applications. REA can be therefore described as a range with a lower and upper limit. In this chapter the target parameter is the P21 (Dershowitz and Herda, 1992), which represents the stopping criterion for 2D stochastic DFN and can be upscaled to its volumetric equivalent, the P32, via calibration procedures to constrain 3D models too. Here is described a new methodology for quantifying the Representative Elementary Area (REA) range, based on the evaluation of three statistical parameters of the P21 value distribution: the mean, the variance, and the distribution shape. Each parameter is assessed using a different statistical test and an associated diagnostic plot. The proposed method combines multiple statistical tests with diagnostic plots, offering a multi-parameter framework for REA estimation. It enables the user to explore the reasons behind different statistical behaviors, enhancing interpretability and reducing the “black box” aspect often associated with statistical tests.

Chapter 1 - An Integrated Workflow for Parametrization of Fracture Network Geometry in Digital Outcrop Models

At the time of thesis submission, this chapter was accepted for publication in Solid Earth as https://doi.org/10.5194/egusphere-2025-1398

Abstract. Mesoscale fractures, with lengths between meters and tens of meters, cannot be effectively characterized in the subsurface, due to limitations of borehole and geophysical datasets. On the other hand, large quantitative structural datasets can be collected on outcrops by combining direct observations and remote sensing (digital outcrop models - DOMs). These data can be used to constrain geometrical models of subsurface fracture networks with the outcrop analogue approach.

In this contribution we present a workflow that leverages DOMs with at least two perpendicular faces and combines multiple types of input data (point cloud, textured surfaces and orthophoto DOMs), to collect a suite of statistical parameters to be used as input in current stochastic 3D DFN (Discrete Fracture Network) models.

Orientation data are collected with a semi-automatic procedure applied to point cloud DOMs of the vertical side of the outcrop to extract 2D polygonal facets. Fracture sets are defined with a clustering procedure and different orientation distributions are fitted and tested with goodness-of-fit tests.

Simultaneously, fracture traces are digitized on textured surface or orthophoto DOMs. Topological parameters are calculated on the digitized fracture network on horizontal and vertical orthomosaics, also considering relationships between fractures and bedding. Trace length and height distributions are estimated with an innovative approach, accounting for the censoring bias with survival/reliability analysis. P_{21} (ratio between total fracture length and sampling area) is measured from traces digitized on the large horizontal outcrop, also allowing for the Representative Elementary Area (REA) to be assessed. Even if the height/length ratio cannot be measured on an outcrop by any means, we attempt to relate heights and lengths under the assumption that the two datasets are correlated, with the longest fractures being also the tallest. We discuss the applicability of our workflow on a large high-quality fractured limestone outcrop in the Murge Plateau near Altamura (Puglia, Italy).

1 Introduction

Fractures exert a fundamental control on the mechanical and hydraulic properties of rock masses, and their relevance extends to multiple applications, including reservoirs of every kind of geofluid (March et al., 2017; Wallace et al., 2021; Wang et al., 2022; Forstner et al., 2025), nuclear waste repositories (Follin et al., 2014; Hadgu et al., 2017),

geology engineering (Eberhardt et al., 2004; Agliardi et al., 2017; Franzosi et al., 2023a, b) and contaminant transport (Medici et al., 2024; Cherubini, 2008). In all these applications, fracture patterns hold great importance as they influence the direction, magnitude, and heterogeneity of fluid flow, the storage volume of reservoirs (Davy et al., 2013; Wang et al., 2022), and rock mass strength.

Fracture networks are complex geological objects composed of all the fractures in a rock mass. Here, the term "fracture" will be used as a general term including both opening-mode or shear fractures (joints, faults, etc.), filled or not (veins, joints, etc.). Broadening the meaning of "fracture" by including other kind of discontinuities, such as deformation/compaction bands, foliations, bedding, pressure solution seams and stylolites, etc., may be useful in some research field or application, such as engineering rock mechanics, geomorphology or hydrogeology (Schultz, 2019; Eppes et al., 2024). Fractures can be classified in sets, i.e. populations of cogenetic discontinuities related to the same deformation phases, kinematics (e.g., joint, normal fault), filling (e.g., quartz vein) and orientation, within statistical variability (Hancock, 1985; Laubach et al., 2019).

Parameter	Fracture network	Fracture set	DOM - Facets	DOM - Traces
Number of sets	*		*	*
Orientation		*	*	
Topology	*	*		*
Size (length/height)		*		*
H/L ratio		*		*
Density/Intensity (1)	*	*		*
Aperture		*		
Spatial organization	*	*		*
Representative Elementary Volume, Area (2)	*	*		*
Roughness		*	* (3)	* (4)
Kinematics		*		
Deformation Mechanism		*		
Filling		*		

Table 1 Summary of the fracture properties needed to quantitatively characterize a fracture network. (1) The P_{xx} system introduced by Dershowitz and Herda (1992) is generally used for density and intensity. (2) The representative elementary volume (REV) can be different for each property and the overall REV of the fracture network can be seen as a combination of REVs for individual properties (e.g., Martinelli et al., 2020). (3) (Candela et al., 2012) (4) (Bistacchi et al., 2011)

The quantitative characterization of fracture networks requires the determination of several geometrical and topological attributes of fractures and their statistical distributions (Table 1). Some of these attributes apply to the individual fracture set (e.g., orientation, length/height distribution) others to the whole fracture network (e.g., topology).

Several factors negatively impact our ability to quantify these parameters, both in the subsurface and in outcrops (e.g. Healy et al., 2017; Laubach et al., 2019; Martinelli et al., 2020):

1) Fractures in the subsurface (e.g. in reservoirs) can only be partially characterized at the mesoscale (meters to tens of meters) using direct techniques. Boreholes provide local information (limited to1D traces in a 3D volume) about the orientation distribution, aperture, fracture abundance (P_{10}, Dershowitz and Herda, 1992) and, if the borehole is properly oriented with respect to the average orientation of a fracture set, 1D spatial arrangement. In contrast, length and height distributions, connectivity and the REA cannot be measured.
2) Geophysical methods can provide continuous 3D information, but with important limitations since (i) fractures are not always associated to a contrast in physical properties that can be imaged with geophysical techniques, and (ii) in any case the spatial resolution of these datasets is limited. For instance, in good quality industrial 3D seismics, fractures smaller than about 200-300m cannot be detected, and, in order to be directly observed, these fractures should be characterized by a displacement that results in a contrast of seismic impedance across the discontinuity. Summing up, only macro-scale faults can be reasonably imaged in seismics, and this induces a biased estimate of volumetric fracture metrics (e.g., Laubach et al., 2019).
3) At the outcrop scale four major biases must be taken into account (Baecher and Lanney, 1978): orientation bias, truncation bias, censoring bias and size bias. The orientation bias stems from the nature of the intersections between the fracture plane and the outcrop surface and to the choice of the sampling dimensionality (e.g., lines, areas or volumes). It influences the representativity of field measurements, and results in downsampling of certain fracture sets with respect to others. The truncation bias imposes a lower boundary to the measured fracture trace length, and it is defined by the smallest feature that is possible to detect. The censoring bias is due to the finite nature of outcrops since the full length of the longest fractures is limited by the outcrop size, and in any case the length of fractures ending outside of the outcrop is not known exactly. The size bias states that larger fractures (i.e. fracture surfaces with a larger area) have a greater probability to intersect the outcrop surface and to be sampled. Another bias, related to layered media, is the under-sampling of fractures shorter than the bed thickness (Ortega and Marrett, 2000). This bias changes the shape of the length distribution, given that only the fracture high enough to abut or crosscut the bedding interface can be systematically sampled. To these major biases it is important to add that the morphological and weathering conditions of the outcrop strongly influence the calculation of parameters like topology, density and intensity. In addition to objective biases related to outcrop geometry or sampling methods, subjective biases introduced by the interpreter should

also be considered (Andrews et al., 2019). In the specific context of automatic feature extraction, it is also important to account for biases inherent to the algorithms themselves and data resolution, including the potential for extracting artifact features.

4) A complete 3D description of the fracture state is only possible in the lab at the centimetric to decametric scale, using non-destructive imaging techniques such as X-Ray Computer Tomography (Agliardi et al., 2014, 2017), which allow measuring volumetric parameters such as P_{33} (fracture porosity, i.e. fracture volume per unit volume), P_{32} (volumetric fracture intensity, i.e. fracture area per unit volume) and P_{30} (volumetric fracture density, i.e. fracture number per unit volume; Dershowitz and Herda, 1992).

The impossibility to directly map or image fractures in the subsurface lead to using continuum representations based on some form of upscaling or homogenization, such as the dual porosity model (Warren and Root, 1963). Alternatively, the Discrete Fracture Network (DFN) approach allows generating stochastic simulations where fractures are simplified as planar polygons in 3D or segments in 2D. In the standard and most widespread approaches to stochastic 3D DFNs, the geometrical properties of each fracture are drawn from parametric length and orientation distributions, and fracture height is generally controlled by a fixed height/length ratio. The simulator generates fractures until a target fracture intensity P_{32} (Dershowitz and Herda, 1992) is reached in the simulation volume (e.g. Move – https://www.petex.com/pe-geology/move-suite/, Petrel –https://www.slb.com/products-and-services/delivering-digital-at-scale/software/petrel-subsurface-software/petrel, FracMan – https://www.wsp.com/en-gl/services/fracman, DFNworks – https://dfnworks.lanl.gov/). Fractures are randomly distributed in the simulation volume according to a Poisson point process, therefore connectivity or any other form of spatial organization cannot be reproduced in these models. More sophisticated approaches have been developed in the last years to try and solve this fundamental limitation (Bonneau et al., 2013; Davy et al., 2013; Bonneau et al., 2016; Bruna et al., 2019; Shakiba et al., 2024; Taty Moukati et al., 2025), but a satisfactory solution has yet to be found, especially in 3D.

Due to the beforementioned limitations in subsurface datasets, input properties for generating stochastic DFNs are often obtained from representative analogues exposed in outcrops that can be characterized, compensating for the information gap at the reservoir scale. The outcrop analogue approach assumes that the detailed information gathered at selected, high-quality rock outcrops can be considered representative of the fracture network properties of deep rock masses that underwent a geological and tectonic evolution that is at least partly comparable. The applicability of an outcrop as an analogue should be evaluated carefully, and some assumptions should be eventually made (Forstner and Laubach, 2022).

This approach relies on the availability of extensive datasets to characterize statistical distributions of the fracture network. In this regard, field survey, intended as physically inspecting and collecting data from outcrops, is a fundamental step in the process of fracture network characterization, because features such as kinematics,

roughness, relative chronology and mineralization/filling can only be gathered during fieldwork. At the same time, even if it is possible, manually collecting massive amounts of data is time consuming on horizontal outcrops, and very difficult in vertical outcrops, where the accessibility is limited (data can only be collected in the portion of the outcrop reachable by the geologist) and depending on the conditions, safety is not guaranteed (e.g. rocks falling from the top of the cliff). To solve this problem, Digital Outcrop Models (DOMs) - high-resolution 3D photorealistic representations of natural outcrops (Bellian et al., 2005; Bistacchi et al., 2022b) have been successfully employed to collect large quantitative structural datasets, overcoming the limitations of classical survey techniques (Sturzenegger and Stead, 2009; Gigli and Casagli, 2011; Sturzenegger et al., 2011; Riquelme et al., 2014, 2015; Bistacchi et al., 2020; Martinelli et al., 2020; Bistacchi et al., 2022a; Storti et al., 2022).

Depending on the outcrop morphological expression, data can be collected from DOMs using either facets - 2D planes interpolated on the DOM, or traces - polylines that are usually digitized in a GIS environment, but sometimes also on a 3D DOM (Bistacchi et al., 2022b). These two types of data carry different but complementary information; however, the methodologies developed in previous contributions by different authors are often based on only one of these kind of data, limiting the number of parameters that can be obtained (Ortega et al., 2006; Boro et al., 2014; Martinelli et al., 2020; Smeraglia et al., 2021).

The scope of this paper is to present a workflow based on statistically rigorous methodologies to characterize a fracture network from the geometrical point of view. The result of such workflow provides a suite of parametrical distributions to be used as input in current stochastic 3D DFN models. The parameters considered here are: The orientation distribution, the length/height distributions, the topological parameters, the fracture areal intensity (P_{21}) and the H/L ratio. We aim at integrating 2D and 3D data sources (point clouds, orthomosaics, DEMs), vertical and horizontal outcrops and facet and trace data to achieve a 3D geometrical parametrization of the fracture network (Sect. 9 and following). The methodologies proposed to estimate each parameter can be applied independently, subject to the type and quality of the outcrop.

The first part of the paper is dedicated to best practices about data acquisition (both ground-based and UAV-based), pre-processing, reconstruction and quality assessment of a photogrammetric model (Sect. 3). Then two separate processing pipelines are presented, depending on the DOM type: (i) semi-automated fracture orientation analysis carried out on point cloud DOMs (Sect. 4.4); and (ii) fracture trace analysis carried out on orthomosaics, allowing to measure topological relationships, length and/or height distributions, P_{21}, and to estimate (subject to assumptions) the H/L ratio distribution (Sect. 5 to 8).

We tested our workflow at an abandoned quarry of the Altamura Limestone Fm. (Puglia, Italy), where both a horizontal pavement and vertical walls provide the opportunity to fully characterize the fracture network in 3D.

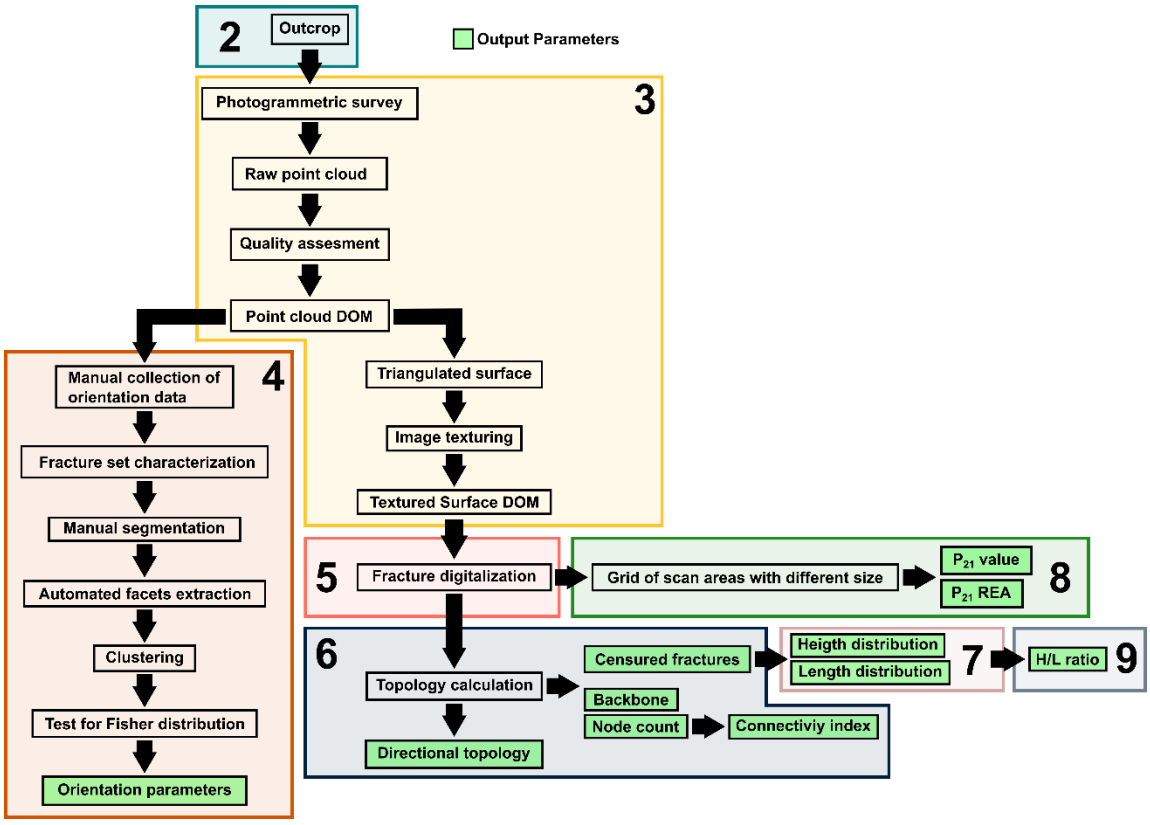


Figure 1 Flow chart of the presented workflow. Numbered boxes define the sections of this contribution in which the respective steps of the workflow are addressed.

2 Selecting an outcrop: the Altamura Limestone at Pontrelli quarry

Outcrops for quantitative fracture survey needs to be carefully selected, in order to satisfy some requirements: a) representativity of the structural and lithological properties of the larger rock volume of interest (e.g., lithological characteristics, structural setting, etc.); b) size large enough to be representative for the structures to be investigated; c) continuous unimpeded exposure; d) optimal orientation with respect to the main fracture sets, to minimize orientation biases (Terzaghi, 1965; Zhang et al., 2002). In this context, it is important to select outcrops that present at least two exposed perpendicular sides (e.g., a vertical cliff and an exposed pavements), natural or artificial (e.g., quarry site), for a full 3D characterization of the fracture network metrics.

Here we consider an abandoned quarry (cava Pontrelli) carved into the fractured limestones of the Apulian platform, in the Murge Plateau near Altamura (Puglia, Italy), in the forebulge of the Southern Apennines fold and thrust belt(Panza et al., 2019). The quarry provides 18.000 m^2 of horizontal pavement and vertical walls with a cumulative width of up to 500 m and up to 6 m in height, where fractures are beautifully exposed thanks to the careful maintenance of the site (Figure 2A, B, C) that is carried out because of thousands of dinosaur footprints that were

discovered by Nicosia et al. (1999). The outcrop, well known and described in previous papers (Panza et al., 2015, 2016, 2019) has been recognized as a suitable analogue for reservoirs in related areas (Zambrano et al., 2016).

The quarry is carved into the shallow marine intertidal limestones of the Calcare di Altamura Formation (Coniacian to Early Campanian, Panza et al., 2016). Limestones are well-stratified light-brown mudstones, with wackestone-packstone layers at the bottom of the beds and sometimes algal laminites in the upper parts. Strata are 20 – 60 cm thick and are organized in thickening upward cycles, some meters thick, bounded at the top by major surfaces of subaerial exposure. Bed interfaces often consist of stylolites having teeth both perpendicular to the folded bedding and tangential (slickolites). The outcrop shows three main fracture sets and a set of "major" structures, that are actually major at the outcrop scale, but negligeable at the regional scale (Figure 2B). Set 1 is the most persistent, it is NW-SE striking and, based on abutting relationships, predates all the other sets (Table 2). It presents both joints and meso-faults with a vertical displacement up to a few cm. Set 2 is also striking NW-SE on average, but with a wider scatter, and it also includes both joints and strike-slip meso-faults. However, structures belonging to Set 2 always abut on those belonging to Set 1, showing that Set 2 is younger (Table 2). Set 3 is NE-SW striking and includes fractures that abut on those belonging to both Set 1 and 2. The trace of these fractures, that are limited by older structures, are relatively short, but are responsible of most of the connectivity of the network (Table 2). Aside from the geometrical characteristics, veins are absent in all of the three fracture sets, as well as fibres on small faults (Set 1 and Set 2).

Even though this outcrop is top-quality in terms of fracture parameterization, due to the areal extension, cleanliness of the pavement, and the association between horizontal and vertical outcrops, some limitations must be still evaluated.

The pavement (Figure 1A, B) presents some no data zones – where no data can be collected at all, due to debris patches or the presence of strong concentrations of non-natural features produced by quarrying activities. Other zones distributed across the pavement are partially affected by non-natural, quarrying-related fractures, but with a careful analysis it is still possible to detect Set 1, while Set 2 and especially Set 3, being composed of smaller fractures, are more difficult to interpret and separate from the ones related to quarrying. Regarding the quarry walls, here we present data on the NW wall (Figure 2C), that is less disturbed by quarrying activities and favourably oriented with respect to Set 1 and 2, while Set 3 is sub-parallel to the wall. The wall is around 6m tall, and according to the stratigraphic analysis proposed by Panza et al., (2016), includes a bed package developed above the quarry pavement, which is one major subaerial exposure surface, while other prominent subaerial exposure surfaces are not detected inside the wall.

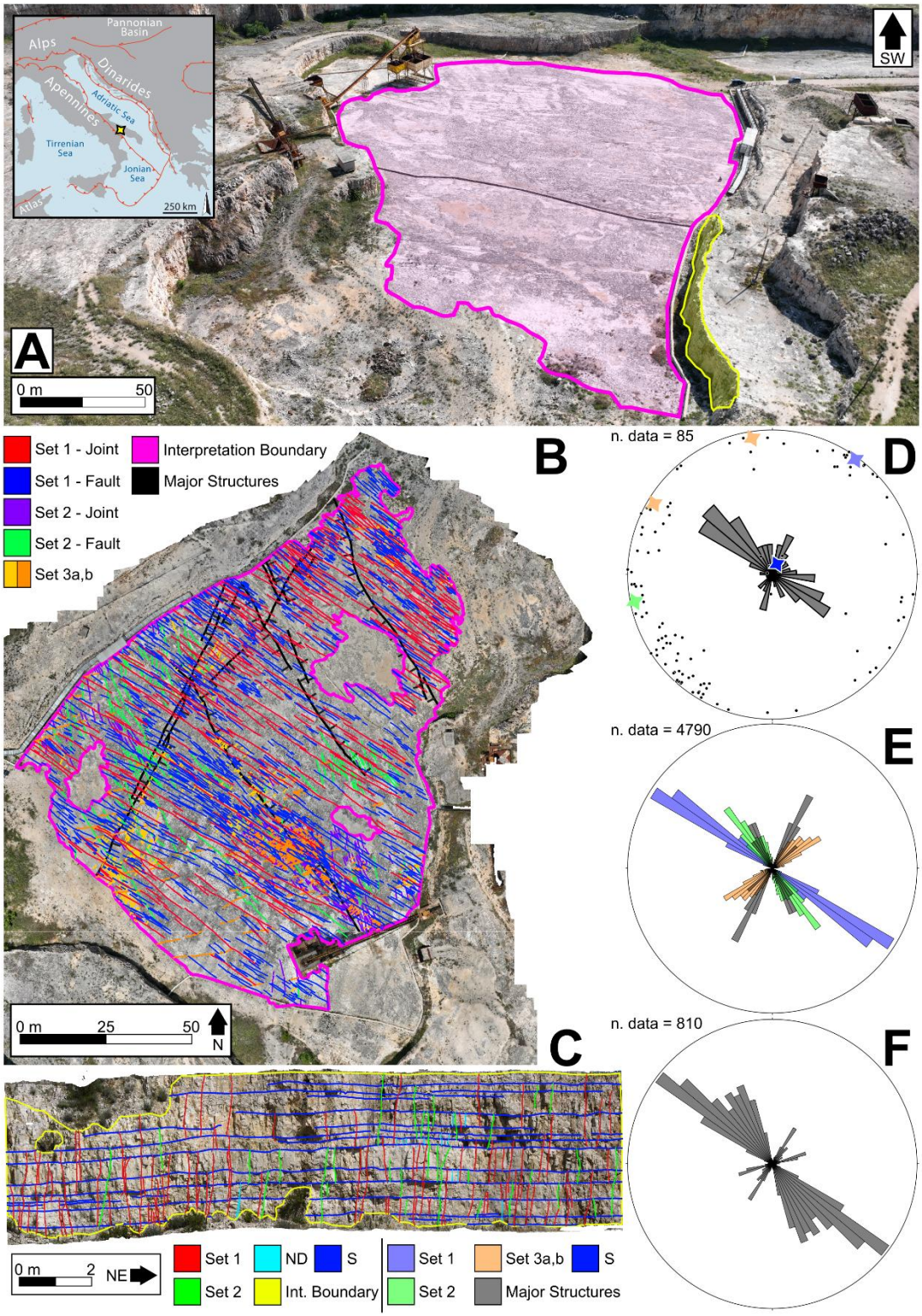


Figure 2 (A) Aerial view of the Pontrelli quarry, Altamura, Italy. The quarry pavement is highlighted in pink, while the analysed quarry wall is highlighted in green. (B) Orthomosaic of the quarry pavement, with digitized fractures and interpretation boundary. (C) Orthomosaic of the interpreted vertical wall, with digitized fractures and interpretation boundary. S represent the stratigraphy. (D) Field data collected along the quarry walls. The stars represent the medoid of each cluster. Each medoid is colorized with the

color of the set according to the legend. (E) Rose diagram of orientation data collected from fracture traces on the orthophoto of the pavement (B). (F) Rose diagram of orientation data collected from fracture facets on the digital outcrop shown in (C).

Fracture set	Structures & kinematics	Average strike	Relative chronology
Set 1	Joints and meso-faults	NW-SE	Abutted by Set 2 & 3
Set 2	Joints and meso-faults	NW-SE	Joints abut on Set 1 & abutted by Set 3. Faults crosscut Set 1 & abutted by Set 3.
Set 3	Joints	SW-NE	Abut on Set 1 and 2

Table 2 Summary of fracture sets characteristics at the Pontrelli quarry.

3 Digital Outcrop Model reconstruction and pre-processing

Once the best exposures have been selected, we must also take care of collecting the best input data in order to create a high-quality DOM, that will greatly facilitate the workflow downstream. This topic was covered extensively by Bistacchi et al. (2022b) and here we just summarize the main requirements in the next paragraphs, always considering the Cava Pontrelli case study.

3.1 Photogrammetric acquisition

Horizontal pavement DOMs have been acquired with a DJI Mavic 3E drone flown with an autonomous flight application (DJI Fly app). The photos were shot perpendicular to the outcrop, with a 70% overlap, both between photos pertaining to a single strip and between adjacent strips. As discussed in Bistacchi et al. (2022b), flights at different altitudes were collected to avoid large-scale distortion in the photogrammetric model, and the minimum altitude of 8m allowed collecting images with a ground resolution of 4 mm/pixel. Georeferencing of these DOMs is based on GPS data collected by the drone and recorded in EXIF data of each photo.

Vertical cliff DOMs have been collected with a Nikon Z7 full-frame mirrorless camera mounted on a tripod with a graduated head, adopting a multiple fan scheme (Bistacchi et al., 2022b), in which every shooting station is evenly spaced by 10° of interstation vision angle, measured targeting a certain point on the outcrop and moving parallel to the outcrop by a distance corresponding to 10°. From each camera locations several photos were shot with a fan pattern, trying to cover the whole outcrop and using different focal lengths, and some shooting stations were collected from a larger distance. This shooting scheme allows (i) avoiding large-scale distortion in the photogrammetric model and (ii) results in an optimal reconstruction of rough outcrop faces, characterized by facets that form a high angle with respect to the main viewpoint. Noteworthy, this kind of survey could be also carried out with a drone, flying and shooting manually, replicating the ground-based multiple fan scheme, but only high-end cinema-grade drones have cameras that can come close to the quality of a high-end full frame DLSR or mirrorless

camera, with significantly higher costs, hence where possible we prefer to use the ground-based technique. The resulting photogrammetric model has a resolution of approximately 2 mm/pixel.

Georeferencing of the terrestrial surveys was simply performed by marking on the outcrop the location of the mirrorless camera shooting stations before carrying out the drone survey. These points were then retrieved from the drone dataset with an accuracy of better than 4 mm (allowed by the high resolution) and used to co-locate the terrestrial dataset in an accurate and perfectly consistent way.

3.2 Point cloud DOM and Textured surface DOM

Regardless of the technique used to acquire the data, DOMs can be rendered, depending on the outcrop morphology and the scope of the work, as point cloud DOMs (PC-DOMs) or textured surface DOMs (TS-DOMs) (Bistacchi et al., 2022b). PC-DOMs, as the name suggests, are dense sets of points, where each point is characterized by XYZ coordinates and an RGB value, and they are the main output of SFM/MVS photogrammetric reconstructions or laser scanning acquisitions. PC-DOMs are particularly suitable to carry on structural interpretations, using specific tools (Thiele et al., 2018), on outcrops where fractures appear as facets of different size and orientation (as in Figure 3). TS-DOMs are derived from PC-DOMs by generating a polygonal mesh from the point cloud and texturing images onto its surface (Tavani et al., 2014; Bistacchi et al., 2015). In this case, the geological and structural interpretation can be carried out in 3D or, as we do in this contribution, with a standard 2D Geographical Information System (GIS) environment (e.g., QGIS).

3.3 Quality of the photogrammetric model

Defining an absolute quality criterion for a point cloud obtained from a photogrammetric survey is not easy, as different kinds of applications have different requirements. In our application scenario, absolute precision is of lesser priority with respect to the relative accuracy within a local reference frame. This can be evaluated in early stages of the photogrammetric processing considering the image reprojection error, measured in pixels, as it directly impacts the relative accuracy of the photogrammetric model as a fraction of its ground resolution (expressed in mm/pixel).

A fundamental requirement in a DOM aimed at structural analysis is that it must be completely free from artifacts (doubled surfaces, distortion, doming), and that noise (isolated points outside the outcrop surface) should be as low as possible. A typical artifact resulting from a low-quality acquisition scheme, that does not include fans or photos collected at variable altitude as discussed above, is the presence of doubled “surfaces”, consisting in layers of duplicated points that do not define univocally the outcrop surface. Bistacchi et al. (2022b), suggested that the best solution is to use a high-quality acquisition scheme, since *a posteriori* solutions do not work or are hugely time-consuming.

Based on our experience, the most important parameter to evaluate the quality of a photogrammetric model for applications in structural geology is the point cloud surface density (SD). By defining a kernel - a sphere of radius R moving in such a way as to being cantered on each point - the point surface density SD can be calculated as the ratio between the number N of points falling in the kernel and the area πR^2 of the largest circle inscribed in the sphere with radius R:

$$SD = \frac{N}{\pi R^2} \quad [1]$$

As an example, in Figure 3, two PC-DOMs of the same vertical outcrop are compared, collected in two different ways to obtain a different SD. The PC-DOM in Figure 3C is reconstructed from more than 400 photos collected as discussed above (fan scheme, with high end camera, Nikon Z7). On the other hand, the PC-DOM in Figure 3D is collected with a smaller dataset (150 photos) collected with a lower quality camera (DJI Mini 3 Pro). The mean *SD* of the PC-DOM shown in Figure 3D is two orders of magnitude higher than the PC-DOM of Figure 3C (298826 vs. 5249 points/m^{-2}), resulting in a much sharper point cloud, from which it is possible to extract more easily, much more structural information.

In conclusion, a good PC-DOM must be free of artifacts, have low noise, and have a high SD on all surfaces of interest, including facets that form a high angle with the outcrop mean plane, which can be properly imaged only if a multiple fans scheme is used.

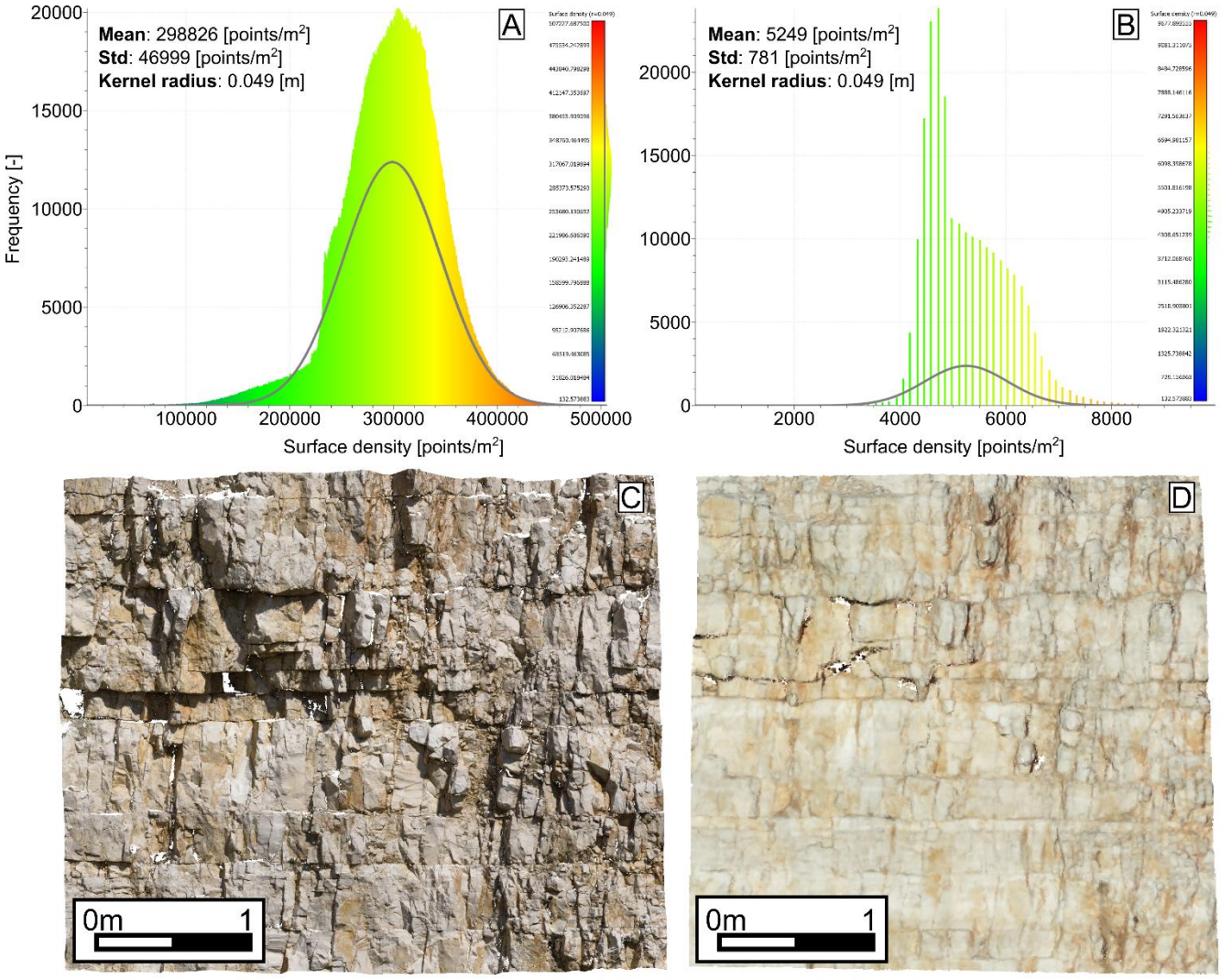


Figure 3 Comparison between two different PC-DOMs of the same vertical outcrop. (A) and (B) frequency distribution of the *SD* measured with a kernel of 0.049m. The parameters are obtained by fitting a Gaussian model. (C) PC-DOM reconstructed from photos collected with a high-resolution full-frame mirrorless camera. (D) PC-DOM reconstructed from photos collected with a commercial drone and a simplified acquisition scheme. Point size has been magnified five times, for visualization reasons.

4 PC-DOM: semi-automated analysis of fracture orientation

The goals of orientation analysis are to measure the attitude of each fracture facet that can be mapped on the DOM and to classify it within a fracture set (i.e. a statistically defined fracture cluster), amongst those identified in the preliminary field survey, or emerging from the clustering analysis (Section 4.3), and finally to obtain statistically validated orientation statistics for each fracture set.

Fractures in PC-DOMs are mainly represented by point patches that are the morphological expression, on the outcrop surface, of fracture surfaces, exposed due to natural (e.g., erosion or rockfall) or anthropic (e.g., excavation) events. Here we propose a semi-automatic workflow to map these patches, based on a first step of manual mapping

of a subset of fracture planes on the PC-DOM. This allows selecting different sets of structures, characterizing their orientation statistics, and assigning them to sets defined in the field. Based on dip and dip direction ranges, the PC-DOM is manually segmented into as many parts as the number of recognized fracture sets. The automatic step consists in the automatic interpolation of 2D planar features from the segmented point cloud, eventually allowing to greatly increase the number of facets included in the analysis, with important benefits for the statistical analysis (Figure 5).

Our workflow can be carried out in CloudCompare (https://www.danielgm.net/cc/), the most used open-source software for point cloud processing (Dewez et al., 2016; Thiele et al., 2018) or in PZero, a new 3D geomodelling application where we are also developing new tools for DOM analysis (github.com/gecos-lab/PZero).

4.1 Orientation parameters for fracture sets

Orientation data are usually recorded in geology using polar coordinates, either as dip and dip direction (dip azimuth) or dip and strike for planar features, or as plunge and trend for axes. In general, any orientation can be represented as a unit vector within a three-dimensional spatial framework (e.g. Mardia and Jupp, 2000), and polar coordinates can be converted into director cosines in a dextral cartesian reference with:

$$L = \sin(dir)\cos(dip) \quad [2]$$

$$M = \cos(dir)\cos(dip) \quad [3]$$

$$N = -\sin(dip) \quad [4]$$

where L is the component in direction East, M is directed towards the North, and N is directed upwards (Borradaile, 2003).

The unit normal vector ($\vec{v} = L, M, N$) is calculated on every point of the point cloud by fitting a plane to the points that fall within a sphere of specified radius, cantered at the point itself. The plane is fitted with a least square fitting algorithm (Dewez et al., 2016). The larger the radius the smoother the normal vectors will result, with the drawback of longer computational times.

With a few exceptions (e.g. bedding with polarity, flow directions), the orientation of geological structures and particularly of deformational features like fractures shows a symmetry where the sense does not bear any geological meaning. In mathematical terms this means that two vectors $\vec{v}$ and $-\vec{v}$ are equivalent in this kind of analysis. This symmetry implies that different conventions can be adopted for the sense of normal vectors, i.e. the geological

convention where normal vectors always point downwards or the photogrammetric convention where they point out of the outcrop.

4.2 Manual orientation mapping

The first step of the workflow is carried out by manually mapping facets, and particularly their attitude, in the PC-DOM with the Compass plugin in CloudCompare (Thiele et al., 2018) or with the facet mapping tool in PZero. These tools behave in the same way and mimic the process of manually collecting attitude data in the field with a geologist compass-clinometer. The fundamental goal of this step of the workflow is to sample each set to define its dip and dip direction range, that will be used in the manual segmentation step. Therefore, we suggest carrying out the mapping with an initial random sampling and then avoiding oversampling the most represented sets (that are generally those favoured by the outcrop orientation bias).

Dip and dip direction values are obtained by fitting a local plane on points selected with a spherical kernel. The kernel radius is qualitatively defined on-the-fly during mapping by the user, based on the dimension of the fracture plane. A too small kernel will result in measurements affected by the roughness of the fracture plane, while a too big kernel will include points pertaining to other surfaces, biasing the orientation value. Orientation data collected in this way are plotted in stereoplots and compared with data collected in the field (Figure 2D), in order to assess whether all field-defined sets are also represented in the digital dataset.

4.3 K-medoid clustering

The precise identification of fracture clusters is fundamental in the following automated segmentation step, where each fracture set corresponds to a cluster of orientation data that can be uniquely defined with a measure of location – a series of measures to locate the fracture cluster in the parameter space (e.g., mean L, M, N), and a measure of concentration or dispersion (Borradaile, 2003). Clustering analysis provides a quantitative answer to both the number of clusters the dataset is composed of, and the parameters of each cluster, given an assumption on the type of distribution.

We apply the K-medoid method to a dataset organized as a table, with n rows corresponding to individual orientation data and three columns corresponding to the three director cosines. K-medoids is a partitional method (Kaufman and Rousseeuw, 1987) aimed at classifying the data into k_m groups, where k_m is the number of fracture set defined in the field, eventually adjusted by the visual inspection of the plotted data . Each group must contain at least one object, and each object must belong to only one group. Partitional methods try to find a suitable partition by separating objects close to each other from objects far away from each other, and how the proximity between objects is calculated determines the specificity of the method. Considering K-medoids in a 3D parameter space (L, M, N), the location parameter is defined by a medoid, i.e. the point belonging to the cluster that minimizes the average

distance in the 3-dimensional space between all the other data belonging to the cluster and the medoid itself (Kaufman and Rousseeuw, 1987, 2005). K-medoids therefore measure distance in an isotropic way in the 3-dimensional space of the dataset. When compared to the more popular K-means approach, K-medoids are more robust and less affected by outliers (Kaufman and Rousseeuw, 1987).

One of the problems that arises with K-medoids and other similar partitional methods lies in the definition of the approach itself, as the number k_m of cluster is imposed by the user, and this can lead to an underestimation or overestimation of the real number of clusters (Kaufman and Rousseeuw, 2005). However, in our application this is not a problem, since the number of fracture sets is iteratively defined starting from an initial guess defined in the field and the clustering algorithm is applied as a validation of that hypothesis, eventually adjusting the number of sets to account for clusters that only surfaced during the statistical analysis.

A second drawback is that in the standard implementation the initial guess for the medoids is chosen randomly, and when different fracture sets show a partial superposition, the clustering algorithm could yield inconsistent and unreliable results. To address this issue, the initial guess can be defined by the interpreter by manually positioning the initial guess for the medoid.

Finally, to avoid the sense ambiguity of orientation unit vectors discussed above, particularly critical for clusters of sub-horizontal vectors that can be mirrored across the stereoplot equator, we have developed a solution based on mirroring all input data. For each input vector $\overrightarrow{v_i}$ we create another vector $-\overrightarrow{v_i}$ with parallel direction and opposite sense (i.e. pointing in the upper hemisphere (Figure 4). Then we perform the K-medoids clustering on this duplicated dataset, extracting $2k_m$ clusters both in the upper and lower hemisphere. Given the symmetry imposed by duplicating the data with mirroring, also the resulting $2k_m$ medoids will be symmetrical, with each medoid in the lower hemisphere having a perfectly symmetrical pair in the upper hemisphere and vice versa, then to conclude the analysis we extract just the k_m clusters with the medoid pointing downwards, in the lower hemisphere (Figure 4).

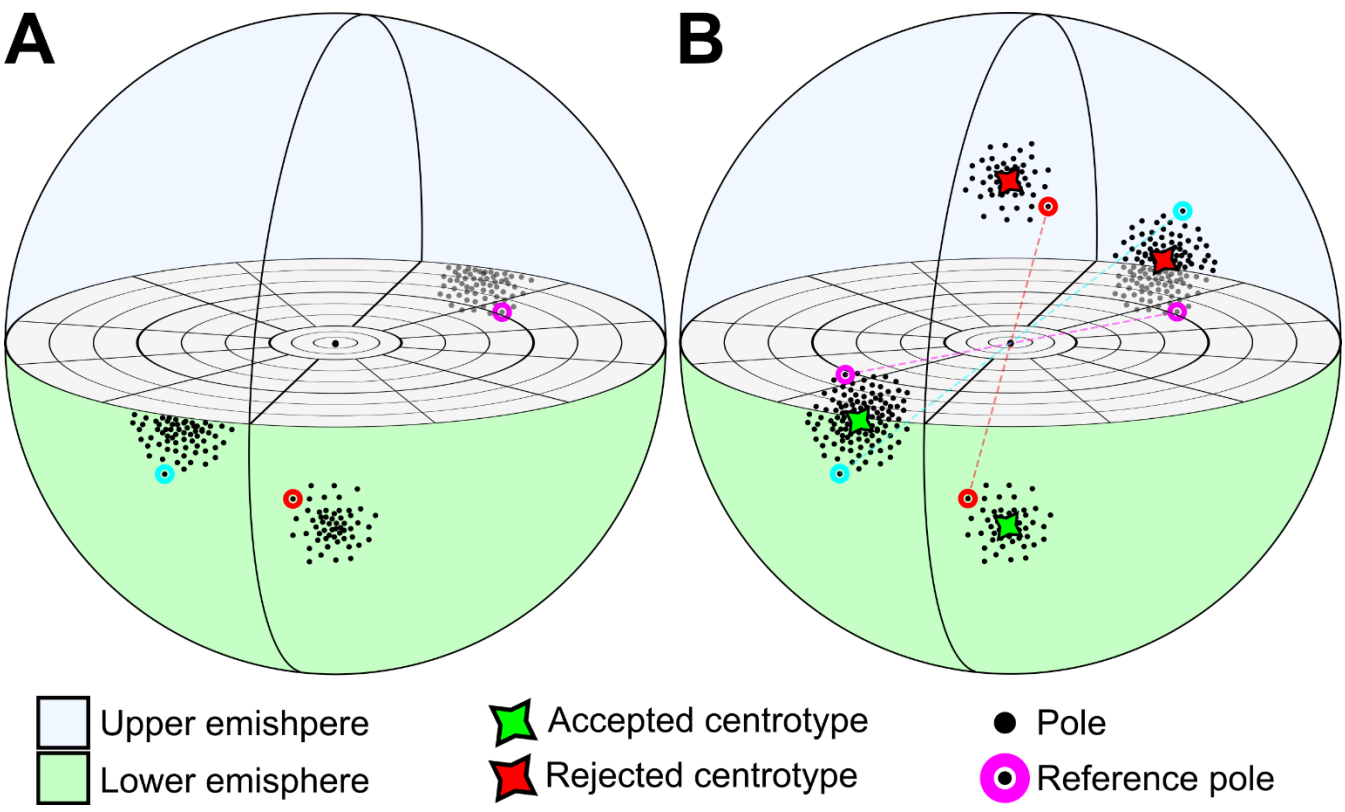


Figure 4 (A) 3D stereoplot of two hypothetical fracture set collected on the field. Light blue and magenta sets are the same set but recognized as two different sets, due to the geological sign convention. (B) Every set is doubled and mirrored with respect to the center of the sphere (B). The clustering algorithm is applied on double the number of the set and only the centrotypes that follow the geological sign convention are kept (B).

4.4 Manual segmentation of PC-DOM and semi-automatic planar feature extraction

Based on orientation statistics and K-medoids clustering, it is now possible to segment the whole point cloud into subsets with normal unit vectors falling within the statistical variability of different fracture sets.

The subsets are composed of isolated clusters of points, with the same orientation, each representing a portion of a separated fracture plane. This greatly improves the automatic extraction of 2D polygonal features during the following steps because every patch of points is isolated from the others, nullifying the risk of merging adjacent clusters into one bigger 2D polygon and avoiding the possibility of generating planes with an averaged orientation between two different point clusters, pertaining to two different fracture sets with different orientations. Another advantage of the manual segmentation is that it is possible to specifically calibrate the algorithm for each fracture set.

The segmented point cloud subsets are then fed to the CloudCompare FACETS plugin (Dewez et al., 2016), and specifically to its fast-marching algorithm (Sethian, 1999), to interpolate a planar polygonal feature per every point

patch that matches the calibrated algorithm parameters. In this context, manual mapping and orientation analysis act as a calibration step in preparation for the final automatic extraction of planar features, that is greatly simplified and results in clean results thanks to the segmentation step, that avoids generating spurious facets.

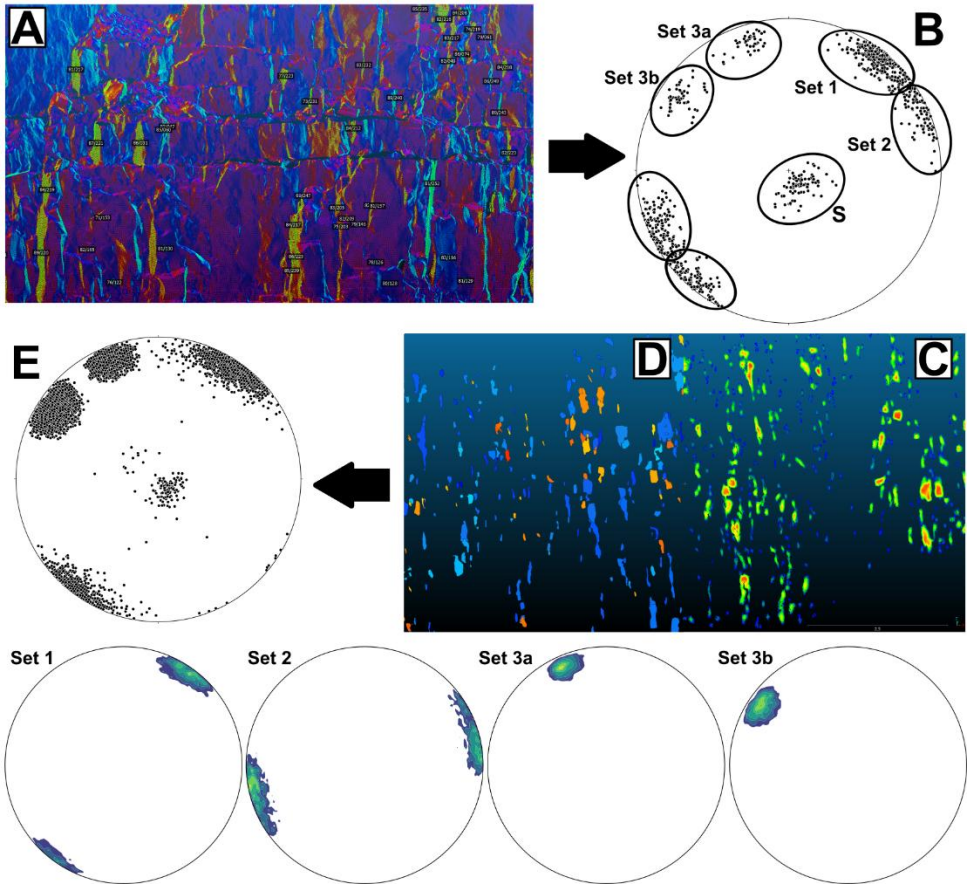


Figure 5 Scheme of the semi-automatic workflow for segmenting the point cloud presented in Section 4. Point cloud colored based on dip direction with a HSV 380 colour scale. (A) Manual data collection on PC-DOM. (B) Manually collected orientation data during the preliminary orientation analysis. Number of data: Set 1 = 351, Set 1 = 256, S = 87, Set 3a = 74, Set 3b = 42 (C) Manual segmentation of the PC-DOM. (D) automatic feature detection with FACETS plugin. (E) Final result of the semi-automatic extraction workflow. Each fracture set is individually shown with contour lines.

Fast marching algorithms are a class of methods developed to track propagating interfaces into a bi-dimensional or three-dimensional space (Sethian, 1999). In the FACETS plugin (Dewez et al., 2016), the fast-marching algorithm is employed to create polygonal planar surfaces interpolating subsets of the point cloud. The algorithm is based on four parameters that need to be calibrated for optimal results, in particular:

- *Octree level* defines the level of systematic recursive subdivision of the point cloud three-dimensional space, defined by its bounding box, in this case, a cube. Every level involves subdividing the box into 8 sub-cubes,

that allow optimizing the definition of the smallest feature we want to detect (i.e. the scale of analysis), with the computational time increasing with the level. No specific strategy exists to calibrate the octree level. As a starting guess we should chose a value that results in cubes with a dimension comparable to the smaller fracture facet expected to be detect. From this value, it is possible to decrease or increase the octree level by one level, visually checking the results. In our experience, increasing too much the octree level does not increase the quality of the analysis but will result in an over-segmentation of the facets and possibly in an increase of noise.

- *Maximum distance* defines a generalization criterion to merge adjacent features. For instance, if maximum distance is set at 68%, at least 68% of the points associated to a facet must have a distance to the facet mean plane that is lower than the standard deviation of the distances from the mean plane fitted from the points defining the facet. In geological terms, this parameter controls the maximum roughness accepted for a plane to be fitted. The maximum distance parameter can be calculated by manually isolating a certain patch of points, representing a fracture plane. The distance from the mean plane of every point can be manually calculated by fitting a mean plane to the point patch. The result is given in the form of a scalar field associated to the point cloud. The mean distance is calculated by fitting a Gaussian model to the frequency histogram of the previously calculated distances.
- *Minimum points per facet* defines the minimum number of points needed to define a facet. The higher the octree level, the smaller this parameter should be, as the dimension of the smallest feature detected decreases and so the related number of points. This parameter can be considered a threshold between what we consider as noise, and what we consider as a proper feature. The minimum points per facet parameter must be tuned according to the average surface density SD of the PC-DOM. The higher the surface density the higher will be number of points in the smaller element produced by the octree subdivision, therefore the higher this parameter can be set.
- *Maximum edge length* is related to the length of the perimetre of the facet. Small values of this parameter impose concave and compact boundaries, while larger values allow for elongated and/or convex boundaries. There is no general rule for the calculation of this parameter, which must be empirically calibrated on a case-by-case basis.

Calibrating all these parameters on the whole point cloud is taxing in terms of computational time, therefore we suggest selecting at least 30 representative facets (Fisher, 1992), in terms of dimension and roughness, for every fracture set, calibrating the algorithm parameters on these facets, and then use this calibration to process the whole PC-DOM. When working on a single facet, the octree level must be set to 0, as an isolated facet is considered as a point cloud on its own (Figure 6).

The result of the feature extraction algorithm is a set of 2D polygonal facets resulting from the interpolation of point patches that met the criteria defined in the calibration, from which orientation parameters will be obtained (Figure 7E).

It is important to remember that facets can be interpolated only on fully exposed planar structures, therefore they represent only the part of the fracture plane that shows a morphological expression. Moreover, based on the calibrated parameters there is the possibility that the interpolation of an exposed surface will result in a combination of facets, and not a single one. All of this to say that data like faces height and surface extension can be useful but should be handled with care if trying to obtain volumetric parameters (P_{30} or P_{32}) or height distributions.

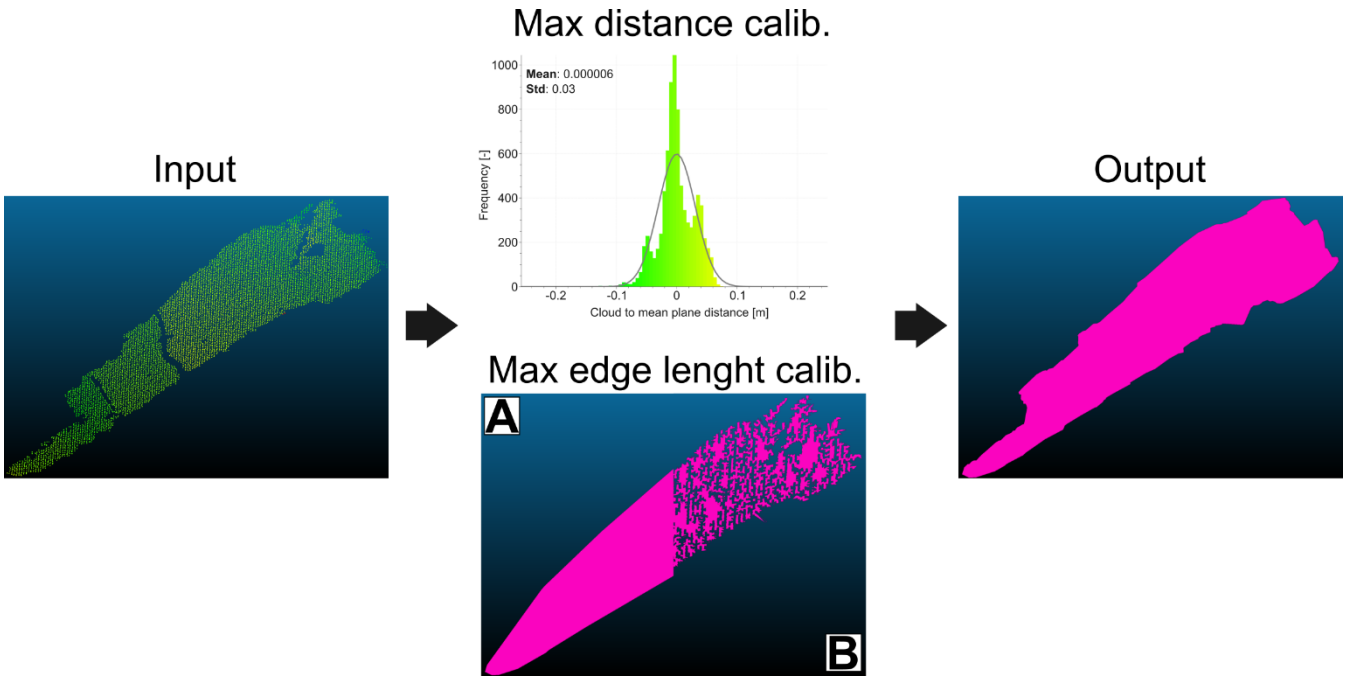


Figure 6 Example of the parameter calibration on a single facet. The max distance is calculated as the mean of the frequency distribution of the distances from the mean plane. (A) Example of a planar feature extraction when the max edge length parameter is too low. (B) When the max edge length is too high the planar feature results in a non-representative polygon. When the parameters are correctly calibrated the output planar feature precisely follows the point cloud border and no point is excluded from the interpolation for a too low max distance value.

4.5 Orientation parameters calculation

In structural studies of fracture networks, it is a common practice to assume that each fracture set shows an orientation distribution described by the unimodal circular symmetric Von Mises-Fisher distribution on a sphere (Fisher, 1953) - a specific declination of the more general Von Mises distribution (Mardia and Jupp, 2000) that, following a common practice in geological applications, will be called “Fisher distribution” in the following.

Even if it is sometimes reasonable to assume that fracture sets follow a distribution with circular symmetry, with the exception of particular situations like radial dikes and fractures formed in a flat layer before the onset of folding (e.g. Mandl, 2005), a statistical test is needed, particularly if the final goal is to use the results of orientation analysis in downstream simulations.

Fisher and Best (1984) proposed a goodness-of-fit test for the Fisher model, starting from a previous graphical test developed by the same authors (Lewis and Fisher, 1982). The poles of the family of n planes that need to be tested (I_j, D_j) with mean dip and direction ($\bar{I}$, $\bar{D}$), are rotated to obtain vectors with new coordinates (I_j', D_j'), with mean dip and dip direction (0,0). The original vectors are then rotated a second time to obtain a new set of vectors (I_j'',D_j''). On these rotated values, the following derived datasets are tested:

- $S_E \equiv \{c_i' = 1 - \cos D_j', 1 \leq j \leq n\}$ [5]
- $S_U \equiv \{I_j', 1 \leq j \leq n\}$ [6]
- $S_N \equiv \left\{Z_i = D_j'' \sqrt{\sin D_j''}, 1 \leq j \leq n\right\}$ [7]

S_E is tested with the Kolmogorov-Smirnov test (Stephens, 1974) against an exponential distribution $E(1/k)$ to check the underlying colatitude distribution (exponentiality test). The Kuiper test (Stephens, 1974) is applied to test S_U against a uniform distribution $U(0,2\pi)$ to check the assumption of rotational symmetry around the mean vector (circularity test). The goodness-of-fit of S_N to a normal distribution $N(0,\sigma^2)$ is tested with the Kolmogorov-Smirnov test (Stephens, 1974) to check against the correlation between colatitude and longitude (normality test) (Figure 7).

Overall, this procedure provides a quantitative way to assess if the dataset can be fit with a Fisher model, allowing the calculation of the mean dip and dip direction of the cluster and the concentration parameter k.

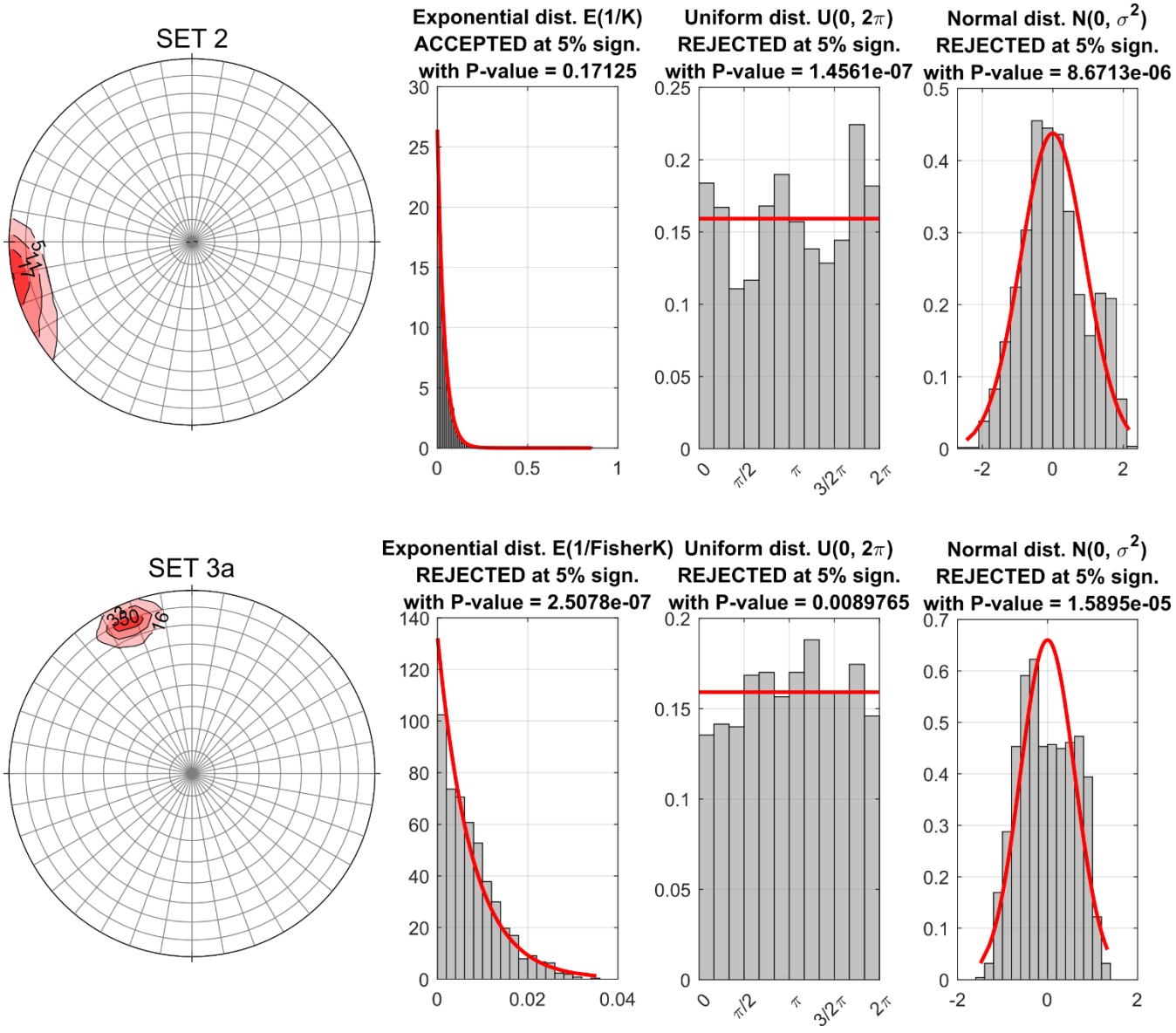


Figure 7 Example of the goodness-of-fit test for Set 2 and Set 3a of the Pontrelli quarry. The hypothesis of sphericity of the data is rejected based on all the tests.

If the Goodness-of-fit test is rejected, it is possible to fit a more general distribution. The Kent distribution is a natural extension of the Fisher distribution as it represents the analogue of a general bivariate normal distribution on a sphere (Kent, 1982). It includes the additional parameter β that describe the ovalness of the distribution, allowing to fit more elliptical clusters (Kent, 1982). No goodness-of-fit tests exist to check if the data follow the Kent distribution. Noteworthy, the Kent model results in orientation distributions that are similar to those off the Bingham-5-parameter distribution (Kent, 1982), but has a different mathematical formula.

5 Enhanced interpretation on orthomosaic and DEM

Collecting orientation data directly from outcrops that lack a noticeable 3D morphological expression (i.e. the facets discussed above) is not possible. At the same time, measuring fracture size and intensity or density (length, height, P_{21}, etc.) from PC-DOMs is not reliable because facets do not correspond to complete fracture surfaces. For these reasons, we digitize fracture traces on TS-DOMs, obtained by projecting and merging the images collected during the photogrammetric survey onto a polygonal mesh or a DEM (in turn interpolated from the PC-DOM). The data extracted from fracture traces are different and complementary to those provided b fracture traces, and only combining both kinds of information the most complete datasets can be extracted from a DOM. The digitization of fracture traces on the vertical TS-DOM is done considering also the corresponding PC-DOM. By integrating TS-DOM and PC-DOM data, each digitized fracture trace can be associated with a best-fit plane derived from the point cloud. This approach enables the assignment of fracture traces to specific fracture sets. Fractures on the vertical wall that could not be reliably linked to a fracture plane were excluded from the digitization process.

The dataset deriving from the interpretation of TS-DOMs is composed of an interpretation boundary (closed polygon) and a series of fracture traces (polylines) attributed to different fracture sets. The interpretation boundary limits the portion of the outcrop where fractures can be detected and digitized. It can include holes to isolate parts of the outcrop, covered by debris or vegetation, that are large enough to hinder the interpretation.

In this case study we were able to obtain orthophotos (see Section 3) of both the sub-horizontal pavement and of sub-vertical walls (Figure 2), and this allowed carrying out the fracture trace digitalization in a 2D GIS environment.

The availability of both RGB images and DEM for the sub-horizontal pavement (from which we can derive slope, aspect and hillshade), allow following structures that may be challenging to detect in RGB images only, due to alteration of the pavement surface, lack of colour contrast or zones damaged by quarrying activities, where longer fractures can still be digitized but are difficult to detect.

Every fracture set is saved in a dedicated file and every characteristic pertaining to a specific set is recorded into an attribute table field. For example, both set 1 and set 2 include meso-faults and joints, this information is stored in an integer field coded as 1 for faults and 0 for joints. At the same time, set 3 is characterized by two main average orientations (Set 3a and Set 3b), but the fractures can be associated by their average length and abutting relationships with other sets. Fractures of Set 3a and 3b are then separated in a specific field.

Precise termination (snapping in GIS jargon) of abutting fracture traces is managed automatically, defining a threshold distance quantified in pixels. In the following sections we discuss how we characterize topological relationships, length/height distribution, H/L ratio and P_{21} from digitized fracture traces.

6 Fracture network topology

Due to the limitations imposed by observing them in outcrop, the topology of fracture networks, which are actually composed of fracture surfaces embedded in a 3D rock volume, is most of the times characterized in 2D, from fractures traces limited and/or connected by nodes (Dershowitz and Einstein, 1988; Barton et al., 1989; Renshaw, 1996; Manzocchi, 2002; Sanderson and Nixon, 2015; Sanderson et al., 2019). Even under this limitation, topology is a fundamental component of fracture network analysis because it is directly related to connectivity, as demonstrated by Sanderson and Nixon (2015).

Topological relationships are also instrumental in calculating unbiased length and height distributions, because topology allows identifying censored fractures by means of B nodes (Benedetti et al., 2025), and this also cascades into the estimation of the H/L ratio.

6.1 Standard topological analysis

From a topological point of view, a fracture network can be seen as a connected set of branches (fracture traces) and nodes (terminations and intersections), delimited by an interpretation boundary (e.g. defined by the natural limits of an outcrop). Six main nodes categories can be defined in a fracture network (Benedetti et al., 2025; Forstner and Laubach, 2022; Nyberg et al., 2018, Figure 8):

- I nodes: fracture trace true tip points;
- Y nodes: abutting relationship;
- X nodes: crosscutting relationship;
- V nodes: perfect coincidence of two tip points belonging to two different fractures - these are theoretically possible, but hard to recognize at the interpretation scale;
- B nodes: boundary nodes, where a fracture trace terminates at the interpretation boundary.
- C nodes: Contingent nodes that can be enabled or not, generating different fracture network configurations, depending on configuration rules defined according to the study objectives and sometimes micro-scale observations (Forstner & Laubach, 2022).

The nature of I, Y, X, V and C nodes is related to the processes that generate the fractures in the first place, but an additional consideration pertains to B nodes (Nyberg et al., 2018, Benedetti et al., 2025), which result from the interaction between the fracture network and the size and shape of the outcrop. This interaction leads to the formation of false tip lines (false I nodes → B nodes) and the censoring of fracture traces. To prevent an underestimation of network connectivity, it is fundamental to exclude B nodes from the calculation of the relative proportions of I, Y, and X nodes.

Nodes classification is based on their topological value (Sanderson et al., 2019), representing the number of branches connected to each node. Specifically, I nodes have a topological value of 1, V nodes have a value of 2, Y nodes have a value of 3, and X nodes have a value of 4. B nodes can be categorized as nodes with a topological value of 3, but one branch originates from the fracture trace while the others come from the interpretation boundary. C nodes assume a different topological value depending on the chosen configuration. If they are enabled, the topological value will be equal to 2 (V node), if they are not enabled, topological value will be equal to 1, and one C node generates two I nodes. This choice heavily impacts length and height distribution calculation as it is controlling the results of the topological analysis. Therefore, the decision about connecting or not fractures through C nodes should be made before running the topological classification.

FracAbility (Benedetti et al., 2025) is an easy-to-use original Python library developed for the quantitative statistical processing of fracture networks. Taking as input a set of polylines for each fracture set and a polygon representing the interpretation boundary, it is possible to obtain:

- IYX ternary diagram: the standard classification when it comes to topology (Figure 8A). In this diagram, the normalized distance from each vertex represents the relative frequency of I, Y and X nodes.
- Connectivity Index (CI): the mean number of connections per line, rendered as contour lines on the IYX ternary diagram (Manzocchi, 2002). As a reference, CI = 3.57 was defined as the critical CI for a constant length uniformly clustered system (Manzocchi, 2002).
- Backbone: the largest cluster of connected fractures in the network.

These results represent different aspects of the degree of connectivity of the fracture network. The ternary diagram and the connectivity index give information about the average connectivity of the network. Only when a fracture abuts on another one it is possible to reduce the number of I nodes in the network, thus moving towards the X and Y nodes vertices in the diagram means increasing, on average, the possibility to form large, connected clusters within the network. However, the presence of clusters of connected fractures with a high connectivity index cannot be unravelled by the simple node count. Backbone extraction addresses this problem by highlighting the most extensively connected cluster. Under the assumption that all fractures are open, it also provides a graphical solution to the percolation threshold problem. Specifically, if the backbone spans two opposite sides of the interpretation boundary, it indicates the presence of a giant connected component, allowing for the establishment of a continuous flow (Haridy et al., 2020). As illustrated in Figure 9, the backbone is characterized by a notable increase in the CI value.

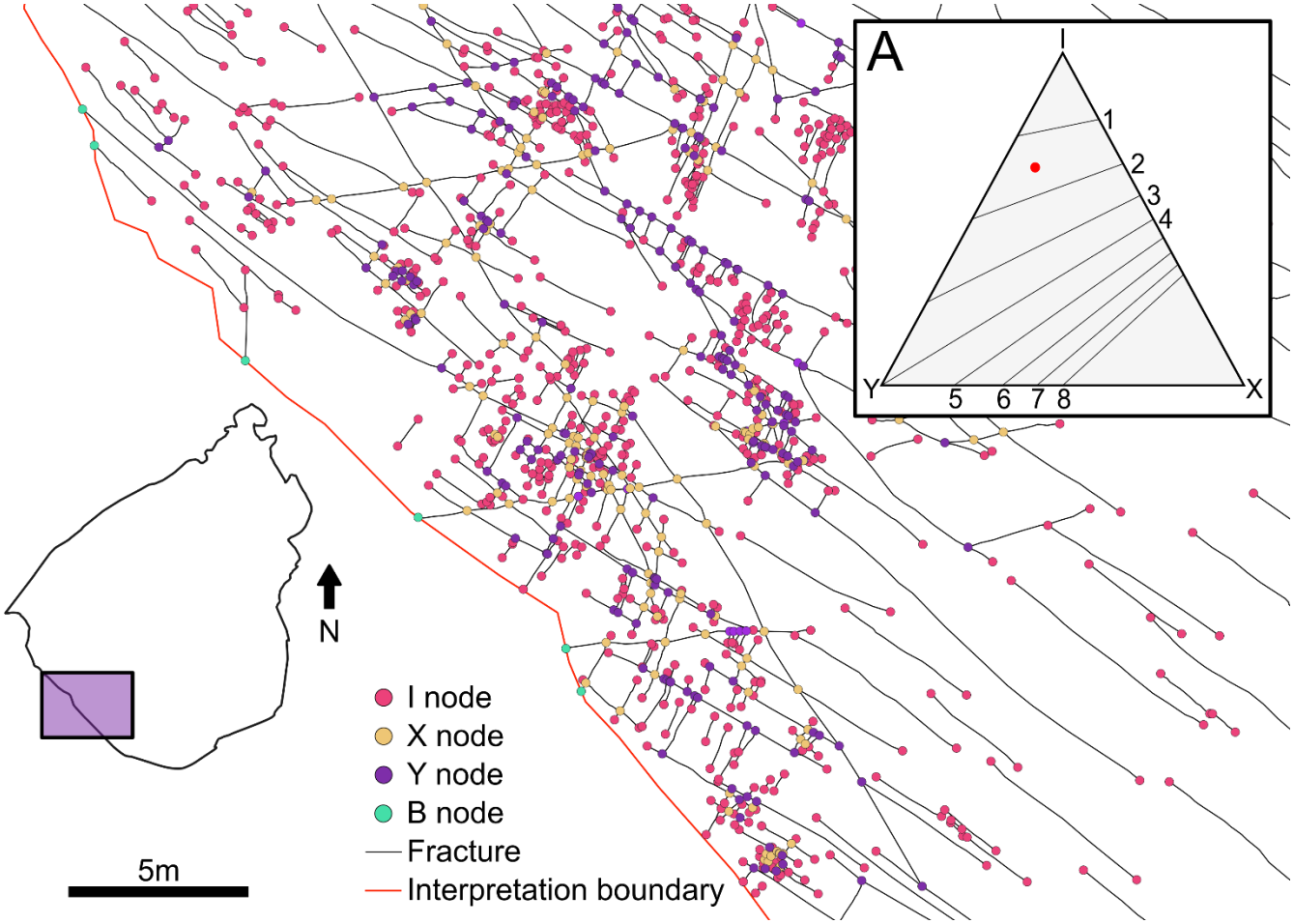


Figure 8 Topological relationship of the Pontrelli quarry fracture network. (A) Associated IYX ternary diagram, the red dot represents the connectivity index (CI).

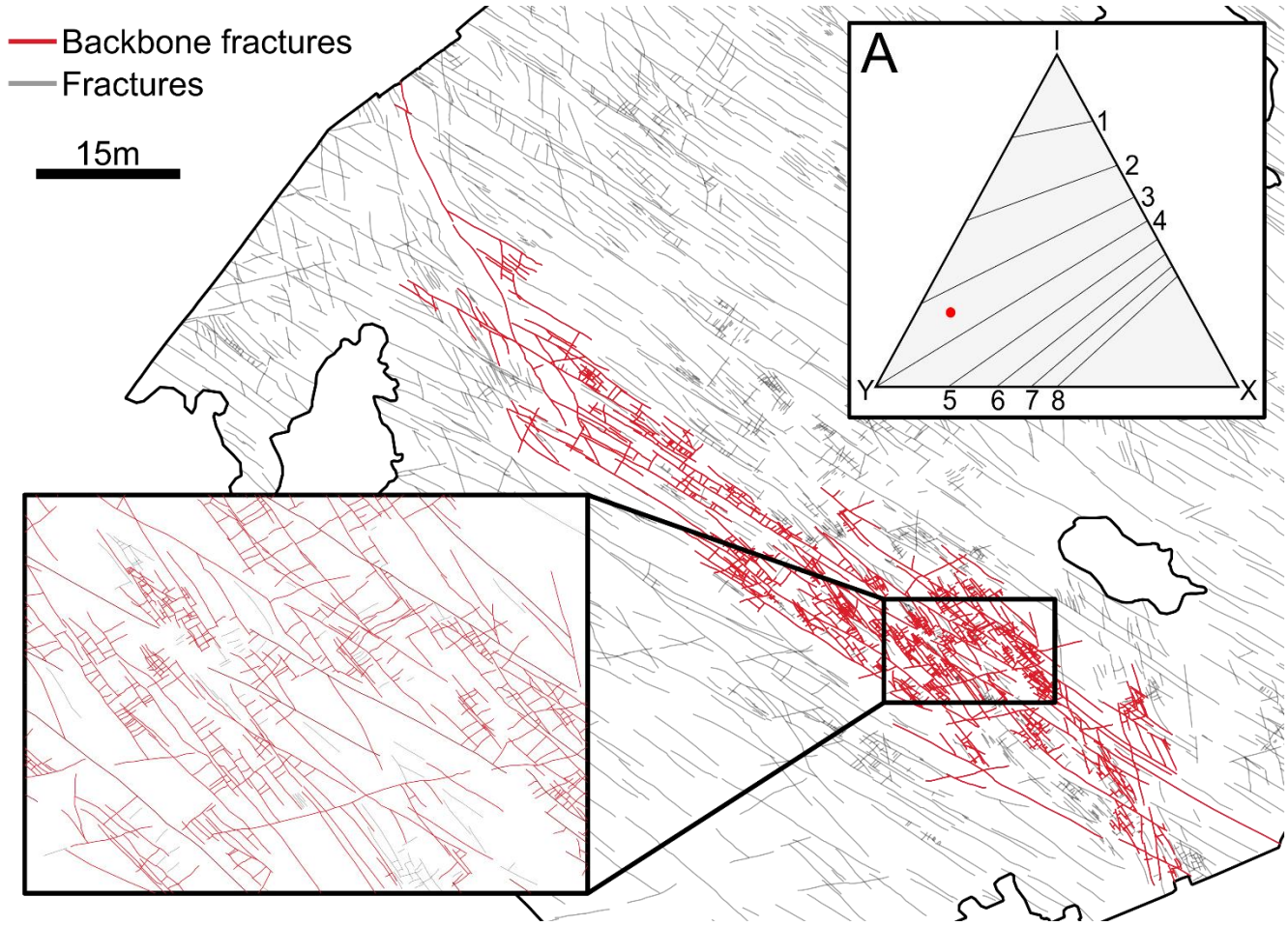


Figure 9 Backbone of the Pontrelli quarry fracture network. The backbone connects two sides of the pavement, indicating the presence of a giant connected component. (A) XYI ternary diagram of the backbone zone, the CI increase form 1.4 for the whole network to 3.5 along the backbone zone.

6.2 Directional Topology

Topological parameters presented in the previous section give a general picture of the fracture network as a whole and are calculated considering the fracture network as a single entity, not considering the geological classification of fractures in different sets (Figure 10A). It is thus impossible to retrieve information about a specific fracture set, for example, how many I, Y and X nodes a certain set have, or how a set is related to another one in terms of crosscutting and abutting relationships. This kind of information can be obtained using what we call "directional topology." In standard topological analysis, nodes store only the topological value. In contrast, in directional topology nodes also contain information about the fracture set (in the case of I-nodes) or sets (in the case of Y- and X-nodes) from which the connected branches originate. This enables a more detailed topological characterization: I-nodes are classified by set, X-nodes are described by the intersecting sets (e.g., an X-node between Set 1 and Set 2), and for Y-nodes, it is possible to determine whether they are generated by Set 1 abutting on Set 2 or vice versa, by counting the number of branches (Fig. 10).

To address this issue, when splitting fractures into branches to calculate topological values, FracAbility stores into the node attributes the set to which every branch belongs and the associated directionality (Figure 10B).

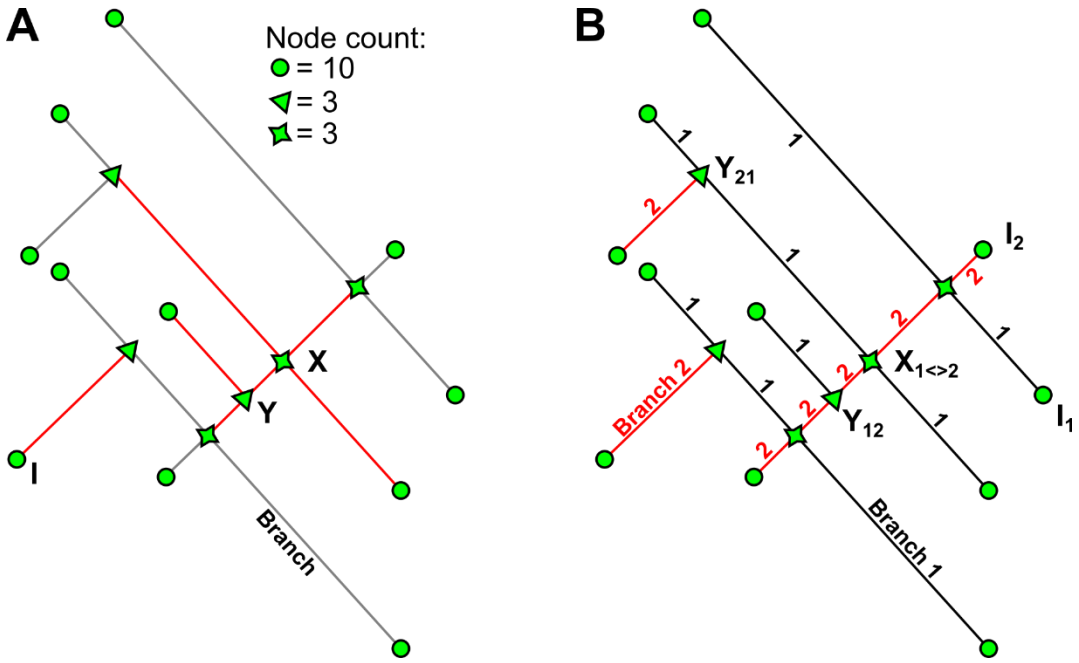


Figure 10 (A) Topological relationships based on the topological value for a fracture network composed of two fracture sets. Highlighted in red, the branches necessary to define one specific node. (B) Topological relationship calculated taking into account the branch origin. X nodes are identified by the presence of two branches for every fracture set, Y nodes are identified by three connected branches. Of the three branches if only one pertain to a specific set it means that it is abutting on the other. I nodes are classified depending on the origin of the connected branch.

The usefulness of directional topology is not only limited to a more advanced description of the topological relationships within the fracture network, but can be also employed to define a quantitative parametrization of relative chronology between fracture sets, and of the stratabound versus non-stratabound nature of fractures.

In a hypothetical case where a fracture network is composed of two fracture sets, without censored terminations, and one of the two sets consistently abuts on the other, the abutting set will only show Y nodes. Thus, dividing the number of nodes by the number of fractures will yield exactly 2. In a real scenario, where the fracture network also includes censored fractures and B nodes, this result will be less than 2 because some of the Y-nodes are masked by censoring. To shield this relationship from censoring, it is necessary not only to subtract the number of censored fracture traces from the total, but also the number of Y-nodes that represent the termination of a censored fracture trace from the total Y-nodes, defining the following relationship:

$$Fracture\ Binding\ Index = \frac{n_{Y\ nodes} - n_{Y\ censored}}{2\,(n_{frac.} - n_{frac.\ censored})} * 100 = 100\% \quad [8]$$

where $n_{Y\ nodes}$ is the number of Y nodes of the abutting set, $n_{Y\ censored}$ is the number of Y nodes associated to a censored fracture (i.e. trace with one B node and one Y node), $n_{fractures}$ is the total number of fractures of the abutting set and $n_{frac.\ censored}$ is the number of censored fractures of the abutting set. The Fracture Binding Index (FBI) ranges from 0, when no fracture from the abutting set abuts on the other set, to 100% when every fracture abuts on another fracture set. 100% of abutting nodes is an asymptotic value, difficult to reach in a natural context, but nonetheless revealing a tendency in this direction would be interesting.

FBI can assume a different meaning depending on the context in which it is applied. In general FBI represents a quantitative way to assess relative chronology. In fact, considering two fracture sets, the one with the higher FBI is interpreted as being younger than the other one since it is consistently abutting. Moreover in vertical outcrops, if we consider the topological relationships between one fracture set and the bedding, FBI represents a quantitative parameter for the quantification of the tendency of a fracture set to be bounded by bedding surfaces.

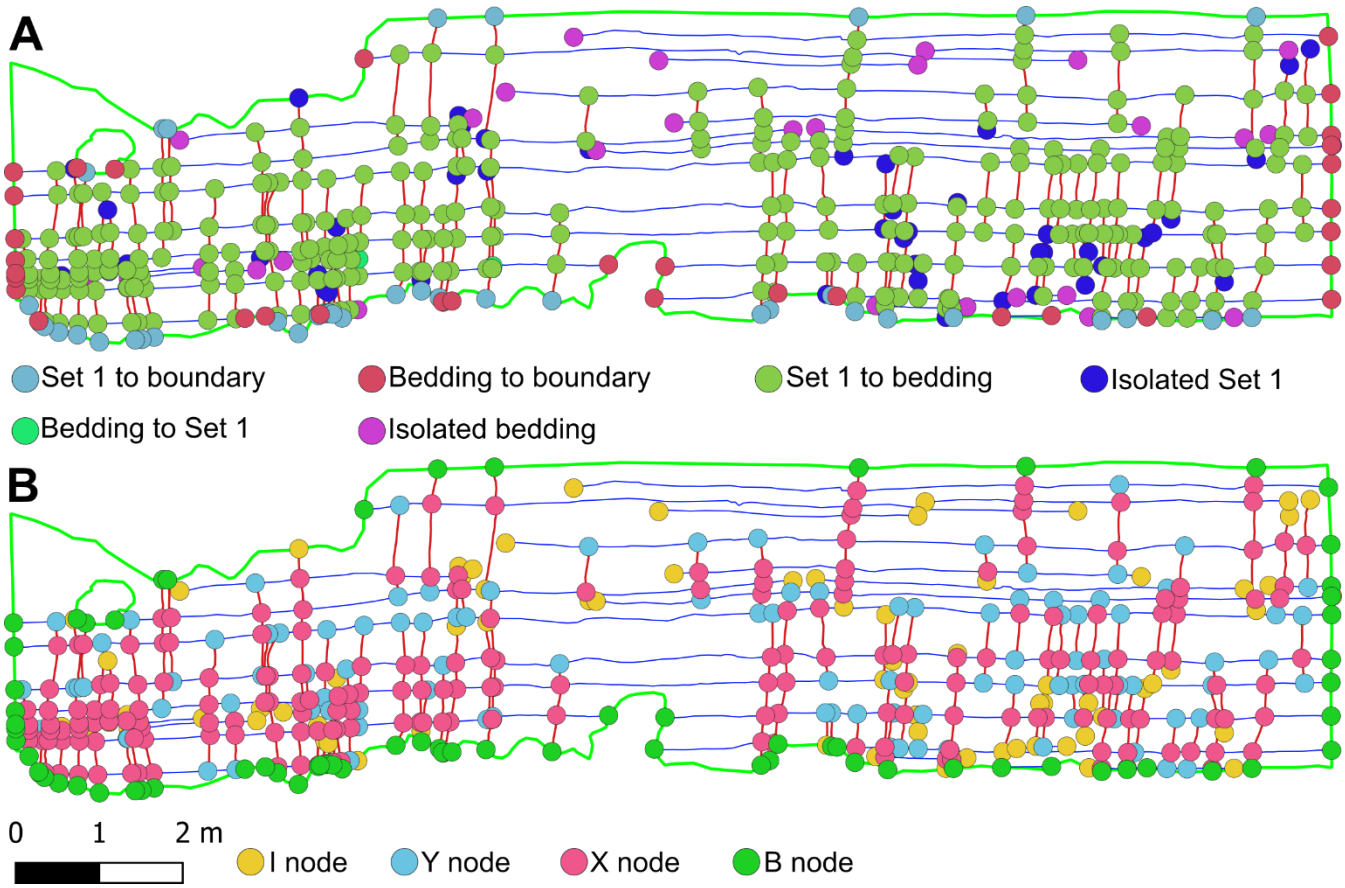


Figure 11 (A) Directional topology applied to Set 1 fractures on Pontrelli vertical wall (B) Standard topological classification of Pontrelli vertical wall. The complete topological characterization is given by the combination of A and B visualizations.

Figure 11 shows the application of directional topology to the Pontrelli vertical wall, considering Set 1 and the bedding (S). Applying the directional topology analysis we obtain:

- $n_{Y\ nodes} = 100$
- $n_{Y\ censored} = 28$
- $n_{\ frac.} = 93$
- $n_{frac.\ censored} = 40$

$$FBI_{1-B} = \frac{100-28}{2\ (93-40)} * 100 = 67\%$$

Therefore Set 1 is stratabound at 67%.

Considering now, for example, the abutting relationships of Set 3 on Set 1 as mapped in the pavement:

- $n_{Y\ nodes\ (3-1)} = 1161$
- $n_{Y\ censored\ (3)} = 4$
- $n_{\ frac.\ (3)} = 1863$
- $n_{frac.\ censored\ (3)} = 12$

$$FBI_{3-1} = \frac{1161-4}{2\ (1863-12)} * 100 = 31.2\%$$

On the contrary:

- $n_{Y\ nodes\ (1-3)} = 183$
- $n_{Y\ censored\ (3)} = 1$

- $n_{frac.\ (3)} = 2003$
- $n_{frac.\ censored\ (3)} = 175$

$$FBI_{1-3} = \frac{183-1}{2\,(2003-175)_{.}} * 100 = 4.9\%$$

Therefore Set 3 abuts with a FBI of 31.2% on Set 1, and on the other hand Set 1 abuts with only a marginal FBI of 4.9% on Set 3, which is considered just an effect of a few digitization errors or local deformational effects.

7 Trace length/height distribution

Defining an unbiased trace length distribution has always been one of the main challenges in rock mass and fracture network characterization. When calculating or estimating trace length parameters, it is possible to distinguish between distribution-dependent (assume a specific probability distribution) and distribution-free methods (population parameters not linked to any specific probability distribution, Mauldon, 1998). On the one hand, distribution-free methods do not rely on any specific assumption about the underlying distribution, but only provide an unbiased estimator of the mean trace length by an indirect correlation (i.e. length is not physically measured, Warburton, 1980; Pahl, 1981; Kulatilake and Wu, 1984; Mauldon, 1998; Zhang and Einstein, 1998; Mauldon et al., 2001; Rohrbaugh Jr. et al., 2002). Not making assumptions about the shape and mathematical form of the distribution could be seen as an advantage, but actually the non-parametrical nature of these approaches implies that it is not possible to obtain statistical parameters of the population such as the standard deviation without imposing further assumptions (Pahl 1981). This makes distribution-free methods unsuitable for modern modelling applications, such as stochastic generation of fracture network, where a fully specified distribution is required. On the other hand, distribution-dependent methods make assumptions on the shape of the underlying trace length distribution, thus constraining their results. Because of this, it is necessary to test how well the chosen distribution fits the data. In the past, this was a strong limitation, due to the biases discussed in Section 1.

Digital outcrops and the increasing computational power make it possible to overcome some problems that previous authors could only consider theoretically from a mathematical and stereological point of view. On the one hand, these new techniques facilitate the acquisition of massive datasets on large sampling windows and successfully tackle the different biases that can be present on an outcrop. On the other hand, the increased computational power makes it possible to calculate the solution to mathematical problems that previously could not be solved due to the lack of a closed form solution (Baecher, 1980).

The orientation bias can be treated by applying areal sampling on outcrops with perpendicular faces. All the fracture sets perpendicular or sub-perpendicular to the horizontal plane are detected on the pavement. If present, fracture sets parallel to the horizontal pavement can be measured on the perpendicular vertical wall, eliminating the issue of under-sampling fracture sets with unfavourable orientations.

The size bias applies to 1D sampling methodologies (scanlines) where longer fractures have a higher probability of being sampled, but this bias does not apply to areal sampling strategies where everything inside the interpretation boundary is sampled. Even fractures much longer than the interpretation boundary are sampled and classified as censored fractures (see below). Areal sampling alone, however, does not account for the possibility of fractures parallel to the outcrop mean plane, and for the under-sampling of fractures shorter than the bed thickness (Ortega and Marrett, 2000). The association between the vertical and horizontal side of the outcrop can partially solve this bias. On the vertical side it is possible to check the presence/absence of a fracture set parallel to the horizontal outcrop surface and the relationship between fractures and the bedding interface. The problem remains for fracture sets parallel to the vertical outcrop mean plane, as the orientation bias hinders the trace mapping. Regarding our case study, we observed that the number of Set 1 fractures on the vertical outcrop roughly matches the number of fractures in the adjacent part of the horizontal outcrop. For Set 2 fractures, we can measure them on the vertical side but they are hidden by artificial fractures related to quarrying activities on the horizontal side. Set 3 is almost parallel to the vertical outcrop configuring the situation in which this bias cannot be evaluated.

Working with DOMs, the truncation bias applies to small fractures that can be truncated by limited DOM resolution. In our case, the resolution of the TS-DOM is around 4 mm/pixel and the smallest digitized fracture is 57 cm. Although the possibility remains that some fractures were missed during the digitization process — including potentially fractures smaller than the identified truncation threshold — the order of magnitude difference between the resolution of the DOM and the smallest recognized fracture is expected to mitigate truncation bias at a fixed scale.

Consequently, the only remaining bias to be treated is the censoring bias, which occurs naturally due to the finite nature of the outcrop or due to the presence of vegetation or debris. In statistics, censoring is a condition where the value of a measurement is partially known. This occurs when some of the data are subject to limitations or restrictions, preventing us from observing the complete information. Censoring can happen for various reasons, and it is a common scenario in statistical analysis (Kaplan and Meier, 1958; Leung et al., 1997; Lawless, 2003). In our case, the fractures that touch the interpretation boundary are objects whose length is partially measured, therefore affected by random censoring (Benedetti et al., 2025).

We thus apply the theory and approaches described in Benedetti et al., (2025) to obtain an unbiased statistical model from censoring of both the length and height distributions. As discussed in Benedetti et al., (2025), this analysis

solves the problems related to single censoring, double censoring, and those related to "holes" within the interpretation boundary, as depicted in Figure 12.

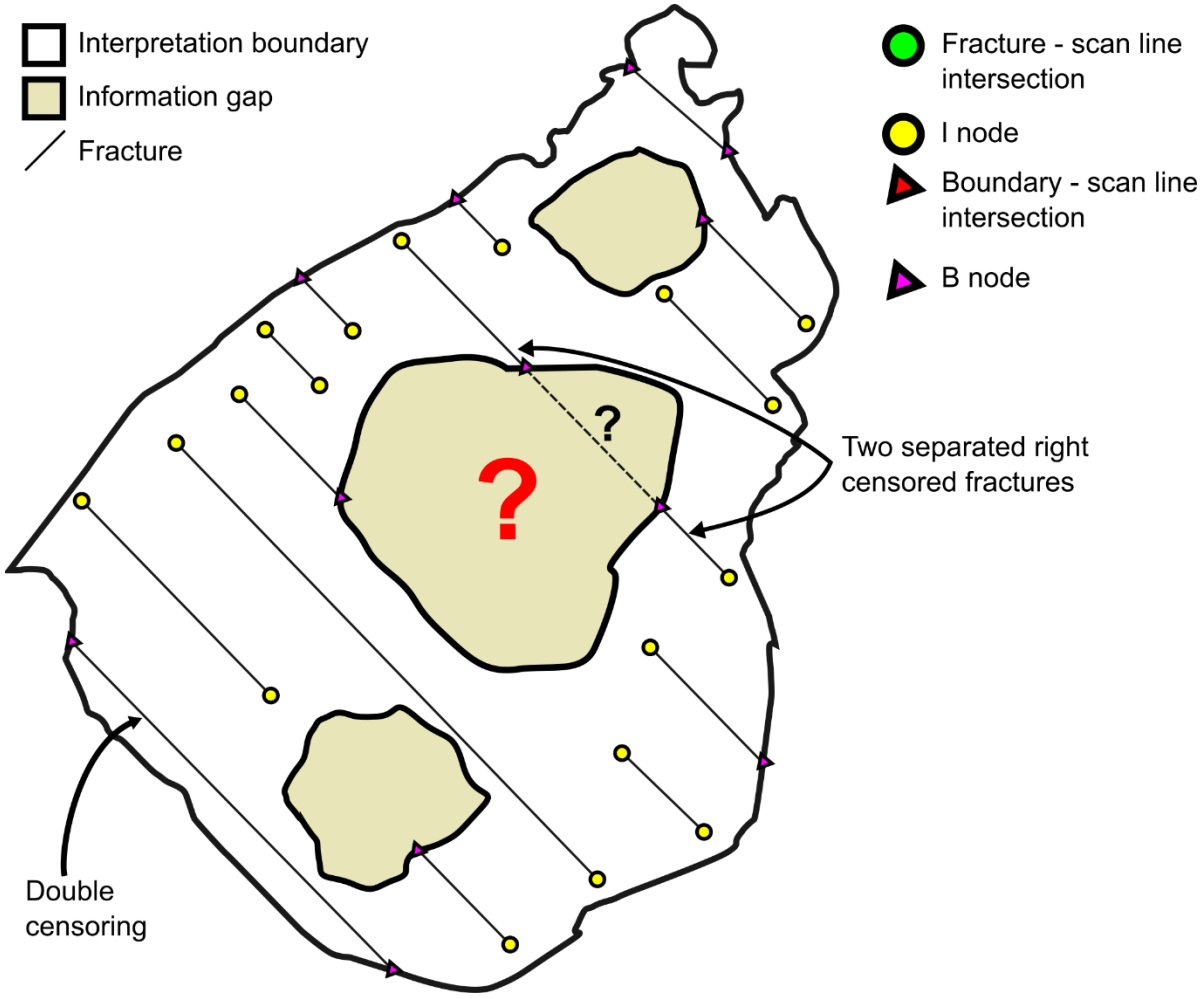


Figure 12 Different cases of censoring in a natural outcrop. The presence of information gaps affects trace length measurements. Double censored fractures are considered a single censored fractures with one of the end nodes coinciding with the interpretation boundary. Fractures that look coplanar across an information gap are considered two separate censored fractures.

Survival analysis parameter estimation is based on optimization algorithms, like the Maximum Likelihood Estimation (MLE), where censoring is taken into account by calculating the likelihood of a censored measurement by using the survival function instead of the probability density function (Benedetti et al., 2025). MLE is a parametric approach that needs a testing phase to validate its results. Working with censored data, without a specified distribution, leads to a situation where none of the standard non-parametric goodness-of-fit test can be applied (Benedetti et al., 2025).

Using the survival analysis approach different hypothesis (statistical models) can be estimated with the censored data. For this case study, we propose to fit the following statistical models: Lognormal, General Gamma, Weibull, Exponential, Gamma, Logistic and Normal. Several statistical distances are calculated between the available empirical data and the fitted model to show which of the proposed models is more representative of the data. We have chosen to use the same distances as (Benedetti et al, 2025) thus using: Kolmogorov-Smirnov distance (DC_n) (maximum distance), the Koziol-Green distance (Ψ_n^2) (sum of squared distances) and the Anderson-Darling distance (AC_n^2) (weighted sum of squared distances), with respect to a uniform distribution U(0,1) (2) Akaike information criterion (Benedetti et al., 2025).

7.1 Distance from U(0,1)

The probability integral transformation theorem (Fisher, 1930) is a fundamental concept in probability and statistics, whose primary application is to transform the values of any random variable into a uniform random variable. The perfect model for fitting a dataset will follow a uniform distribution between 0 and 1, meaning a perfect correspondence between the empirical data and the theoretical distribution, and other models that are close to the uniform U(0,1) with a small deviation will be suitable to describe the data (Figure 13). Therefore, the purpose of the probability integral transformation is to normalize distributions in order to be able to compare deviations on a common ground.

Table 3 shows the rankings based on the different normalized distances for Set 1 fractures. The lognormal distribution and the general gamma distribution rank respectively first and second in all the 3 rankings. In contrast, the logistic and the normal distribution are not suitable for our data. With intermediate rankings, the different meaning of the various distances can be appreciated; for example: the Weibull distribution shows a smaller maximum distance (DC_n) with respect to the exponential and the gamma distribution.

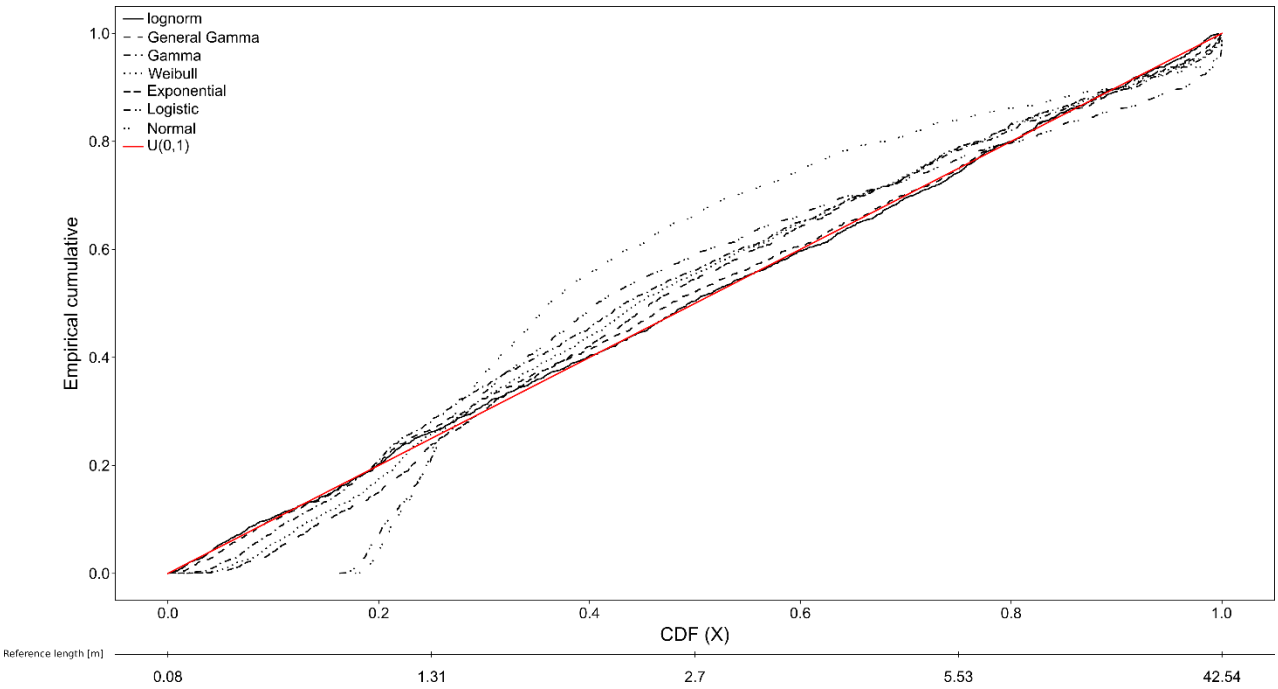


Figure 13 PIT visualization for Set 1 fractures.

Name	DC_n rank	Ψ_n^2 rank	AC_n^2 rank	Mean rank
Lognormal	1	1	1	1
General Gamma	2	2	2	2
Weibull	3	4	3	3.33
Exponential	4	3	5	4
Gamma	5	5	4	4.67
Logistic	6	6	6	6
Normal	7	7	7	7

Table 3 Ranking based on the Kolmogorov-Smirnoff distance, Koziol-Green distance and Anderson-Darling distance for Set 1 trace length data.

7.2 Akaike information criterion

The Akaike information criterion (AIC), is a criterion to rank models from the best to worst based on the empirical data (Akaike, 1974; Burnham and Anderson, 2004). AIC is designed to identifying the so called MAICE (Minimum

information theoretic criterion - AIC - Estimate Akaike, 1974), as the model that gives the minimum of AIC, defined as:

$$AIC = 2k - 2ln(L_{max}), \quad [9]$$

where:

- k: number of model parameters
- L_{max}: maximized value of the likelihood function

Therefore, the Akaike Information Criterion favours parsimony, preferring models with fewer parameters that still adequately explain the data (Akaike, 1974).

Associated to AIC, other important parameters, termed as the Akaike weights (w_i) , give the probability that a certain model is the best model for a given dataset, (Burnham and Anderson, 2004). Their sum is equal to 1 and should be interpreted as weights of evidence, hence the higher their value, the higher the probability that a certain model, in the pool of the selected models, is the best for our data (Benedetti et al., 2025).

The main advantage of this method lies in the use of the maximum value of the likelihood function as input, effectively ranking models and their relative parameters taking into account censored data.

Rank	Distribution name	AIC	w_i
1	Lognormal	7514.617939	0.9912701876
2	General Gamma	7524.082426	0.0087298124
3	Gamma	7636.568978	0
4	Weibull	7653.734152	0
5	Exponential	7655.821815	0
6	Logistic	8804.536054	0
7	Normal	9246.922169	0

Table 4 Ranking based on the Akaike information criterion for Set 1 trace length data.

Just looking at AIC values (Tab.4), and at the various distances measured in the previous section, it seems that even if the lognormal distribution prevails, the three-parameter general gamma distribution remains a valid model for

our data. Akaike weights clarify this situation by showing that the lognormal model is much more powerful than the gamma model at describing our data. At the same time, all other models have Akaike weight equal to 0, which means that they are completely unsuitable compared to the lognormal and the General gamma model.

8 Fracture areal intensity (P_{21})

Fracture areal intensity is defined as the ratio between the total sum of fracture trace lengths and the sampling area (Dershowitz and Herda, 1992; Mauldon et al., 2001). This parameter is very important since its volumetric equivalent P_{32} (total fracture area in unit volume) is used as a stopping criterion in stochastic DFN modelling, meaning that the stochastic generation of fractures will stop when the target intensity is reached, and P_{32} can be obtained from P_{21} via a calibration procedure (Staub et al., 2002). Given the heterogeneous distribution of fractures in natural outcrops, the characterization of this value cannot be separated from the concept of Representative Elementary Volume (REV) or Area (REA) (Bear, 1975). In outcrop studies, REA is the area above which a certain parameter value becomes independent from the position and scan area size with which it is calculated, and can thus be used to constrain larger models.

To perform this analysis while limiting the orientation bias, the outcrop surface was covered with hexagonal grids whose edge length increase gradually, ranging from 1m to 26m in the Pontrelli quarry case study. Only whole hexagons are considered and data are plotted using the graphical boxplot method proposed by Tukey, 1977 (Figure 14A). The lower threshold of REA can be defined as the minimum hexagon area for which no significant difference is detected between the mean and standard deviation of P_{21} obtained at that area and at the next step.

To quantitatively measure the significance of this difference, statistical techniques like ANOVA, used to compare the mean of different populations, can be used in theory (Stahle and Wold, 1989; Moder, 2010). However, ANOVA is based on three assumptions: (i) hypothesis of normality, (ii) homogeneity of variances, (iii) independence between samples (Moder, 2010). In our case, P_{21} samples collected with smaller scan areas are clearly asymmetrical (from 1m to 5m), while P_{21} samples collected with larger scan areas tend to be more symmetrical (Figure 14A). Consequently, variance is inhomogeneous through the dataset, leading to an increase of type 1 errors (Moder, 2010). This problem is enhanced by the fact that the sample size is unequal and decays as the scan area edge length increases due to the finite size of the outcrop (the larger the outcrop, the smaller the number of scan areas). For these reasons, ANOVA and similar tests cannot be applied, and we opted for a qualitative approach based on the difference between the interquartile range (deltaIQR) of two subsequential P_{21} samples. With this approach, REA is reached when deltaIQR stabilizes around 0 (Figure 14B). To account for outliers, that are not included in the IQR, we also consider the range between the whiskers calculated as the difference between the upper whisker length (Q3 +1.5IQR) and the lower whisker length (Q1 - 1.5IQR) (Figure 14C). In both cases, the REA corresponds to a plateau, which, in our case study, is between 5m and 12m of scan-area edge length. Beyond 12m, the representativity is

compromised by the sample size being too small (<15 hexagons). P_{21} REA can be safely calculated only for Set 1 fractures, because, as highlighted in Section 2, in certain areas only Set 1 fractures can be digitized, as Set 2 and Set 3 are obscured by the quarrying related fractures.

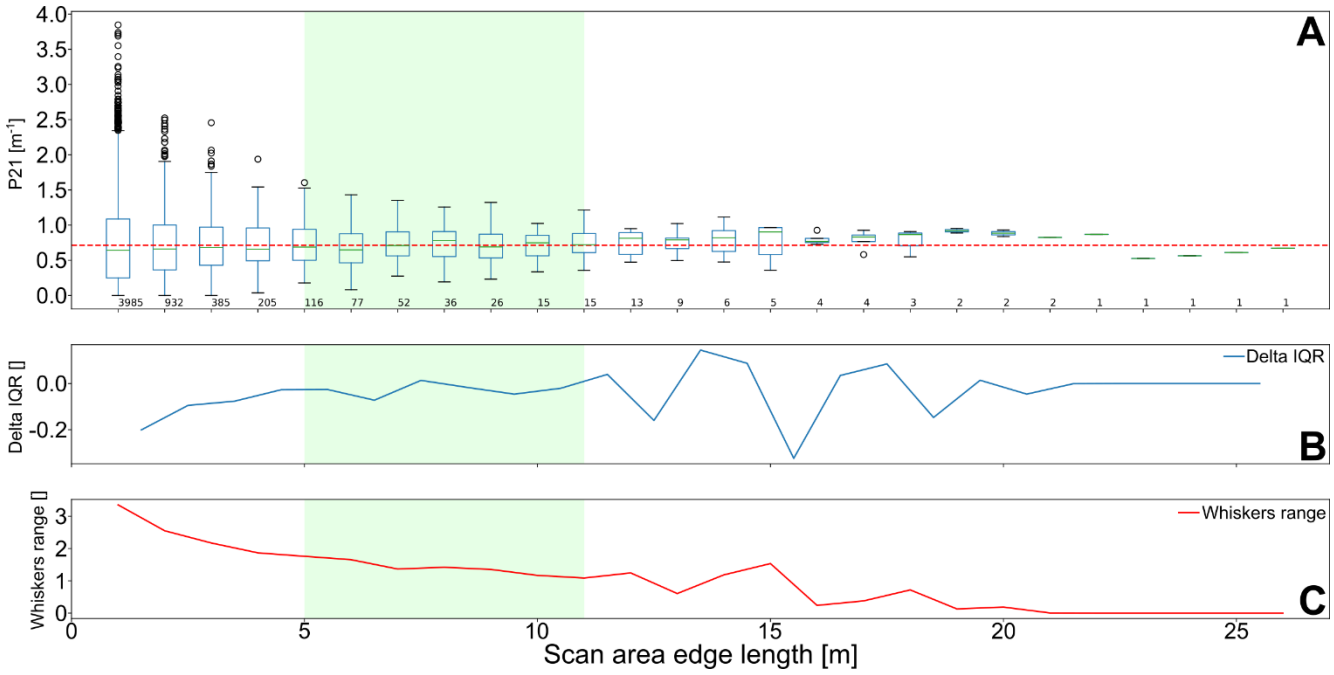


Figure 14 Representative Elementary Area analysis on Pontrelli quarry Set 1. (A) Boxplot of P_{21} data collected with increasing scan area size. Red dashed line: P_{21} calculated on the whole outcrop, with the interpretation boundary as scan area. The green box identifies REA range. Small number under the boxplot: sample numerosity. (B) Delta between IQR of two subsequential P_{21} samples. (C) Range between upper and lower whiskers for each P_{21} sample.

9 Height/Length ratio

The H/L ratio between height of fractures, measured along the dip direction, and their length, measured along strike, is the object of an extensive literature and is believed to span from 1:2 (Odling, 1997; Panza et al., 2015; Giuffrida et al., 2020), to 1:4 (Panza et al., 2015) or even 1:5 (Boro et al., 2014; Smeraglia et al., 2021), depending on different mechanical hypotheses, but could be probably even more variable when conceiving the possible combinations of crosscutting/abutting relationships between fracture sets and with bedding are considered.

H/L ratio is applied, in association with the length distribution, in most commercial and open-source 3D stochastic DFNs to model the geometry of the discontinuities, often represented as rectangular or elliptical surfaces. Therefore, the H/L ratio is directly correlated to the shape of the fracture planes generated by the DFN and controls the switch in dimensionality between 2D and 3D models. Unfortunately, the H/L ratio cannot be directly measured in outcrops due to the impossibility to map the full extension of fracture surfaces (but only their traces or partial facets). In the

studied outcrop, the availability of both height and length data allows us to make at least some realistic and transparent assumption on the H/L ratio based on a correlation of length and height distributions.

We assume is that traces mapped on the pavement (i.e. lengths) and on the wall (i.e. heights) can be associated in ordered pairs from the shortest to the longest. Making some assumption of this kind is unavoidable since there it is impossible to directly observe the correspondence between horizontal and vertical traces. It should be emphasized that this criterion is not unique, and other relationships can be established between length and height data (e.g., random association), but this criterion seems reasonable in terms of fracture mechanics.

To test our hypothesis, a hundred values of length and height are randomly sampled from the statistical distributions of length and height, ordered from smallest to largest and associated in pairs, and the H/L ratio is obtained with linear regression (Figure 16A).

In our case study the height of Set 1 fractures should be limited by the height of the bed package, but the random sampling of the height distribution, that is a lognormal distribution, also generates, although with decreasing probability, fractures that are much higher. This is why in Figure 15 we limited the linear regression to height values smaller than 6m (height of bed package, Panza et al., 2016). One thousand realizations are made to account for the variability of the random sampling and the arithmetic mean of the regression line is taken as the representative H/L ratio (Figure 15B).

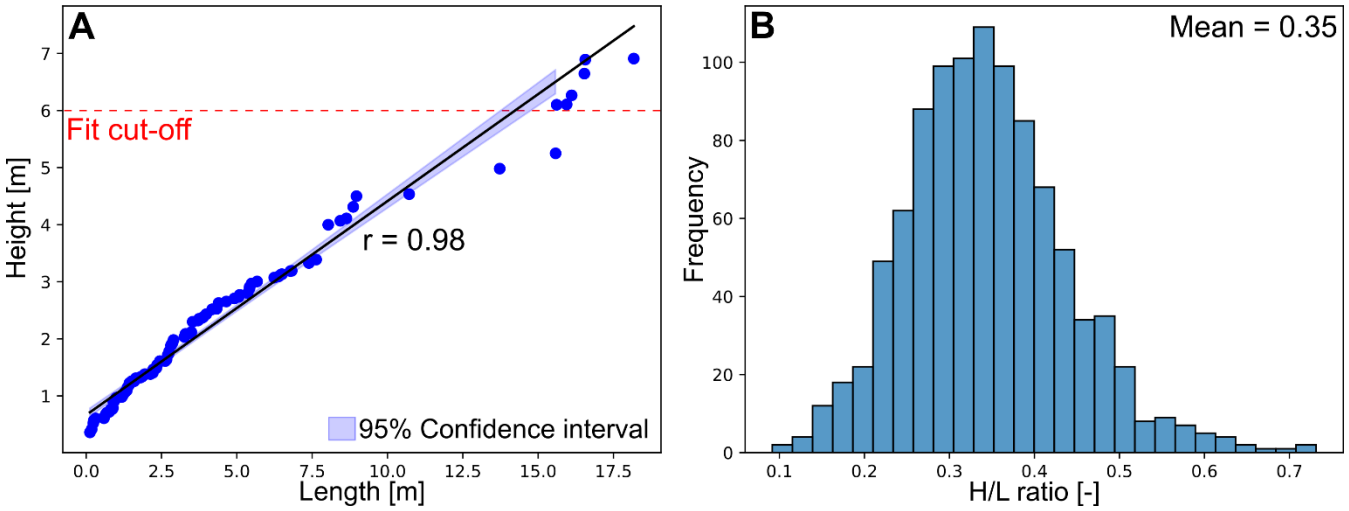


Figure 15 H/L ratio calculated for Set 1 fractures. (A) Example of one realization, where hundred values are sampled from the height and length distributions. r: Pearson correlation coefficient. (B) Frequency histogram of H/L ratios calculated from thousand realizations.

10 Summary of the results: fracture network characteristics at the Pontrelli quarry

In this section, the fracture network of the Pontrelli quarry is described as well as the results and limitations of the parametrization obtained with our workflow. Field observations show that Set 1 is the most prominent set in the outcrop, its fracture traces are homogeneously distributed across the pavement, and it can be detected even in areas damaged by quarrying. From field and DOM observations, all other fracture sets abut or crosscut Set 1, therefore pinpointing that this set is the oldest. This conclusion is also supported by the topological, length and height analyses. Set 1 shows a lognormal distribution of lengths with the highest average and maximum, in agreement with a condition in which fractures can propagate freely, in absence of a mechanical compartmentalization defined by previously developed fracture sets, bedding aside (Ackermann and Schlische, 1997). Set 1 also shows a negligible percentage of abutting relationships with Set 2 (2.7%) and Set 3 (4.9%). This marginal number of unlikely relationships, in a geological context in which Set 2 and Set 3 postdate Set 1, can be explained considering a limited reactivation of Set 1 fractures in more recent tectonic phases. On the contrary, both Set 2 and Set 3 exhibit a significant number of abutting relationships against Set 1, respectively 21.41% and 31.2%.

The favourable orientation of the vertical wall supports the collection of a consistent statistical sample of height measurements, enabling reliable fit to the height distribution. However, the height of the vertical wall is barely sufficient to get a complete observation window on a bed package. This limits the representativity of the assumption on the stratabound nature of Set 1. Our interpretation relies on the results from the directional topology analysis between Set 1 and the bedding. Since more than 67% of the observed Set 1 fractures abut on bedding surfaces, we believe that it should be vertically confined by the height of the bed package. In relation to classical height pattern classification schemes (Hooker et al., 2013), Set 1 falls between the perfectly bed bounded and the top bounded classes, given that even if the majority of the fracture about on the bedding, some fractures (33% of the non-censored fractures) end between two bedding surfaces. The H/L ratio for set 1 is calculated as discussed above from the trace height and length distributions, under the assumption that height and length values are associated in pairs from smallest to largest. This assumption is supported by the strong linear correlation between the two sets of values (Figure 15) and the resulting mean H/L ratio is 0.345.

Set 1							
	Test result	**Mean dip dir.**	**Mean dip**	**Fisher K**	**Kent K**	**Kent β**	**n. data**
Orientation	rejected	213.94	86.3	36.59	39.59	5.42	1475

	Set 2	Set 3	S (Bedding)		
FBI	2.70%	4.9%	67%		
	Dist. Type	**Mean**	**Standard dev.**	**n. data**	
Length distribution	Lognormal	4.49	6.03	2014	
Height distribution	Lognormal	2.09	1.69	93	
	Value				
H/L ratio	0.345				
		REA			
P_{21}	0.7 m^{-1}	5-12 m			

Table 5 Result summary for Set 1 fractures.

Set 2 mean length is intermediate between Set 1 and Set 3 and the topological relationships exhibit a higher occurrence of Y nodes against Set 1 with respect to Set 3, suggesting that Set 2 is the second oldest. The orientation of the vertical wall still allows for the collection of both faces and traces of Set 2 fractures, albeit there is less trace height data than for Set 1, resulting in a less constrained height distribution. This is also reflected in the H/L ratio calculation, which should be applied with greater caution to stochastic modelling.

Unlike Set 1, Set 2 fracture traces in the northern part of the outcrop are masked by the non-systematic fractures generated by quarrying activities. Face and trace data collected on the wall prove however that set 2 fractures are developed in this sector of the quarry. The incomplete sampling of fracture traces across the pavement introduces a strong bias that hinders the calculation of the P_{21} for Set 2, because without a sufficiently wide sampling area, the REA calculation is not representative. Therefore, if this fracture set is modeled with a stochastic approach, some assumption on the REA must be made.

Most of Set 1 and set 2 fractures on the vertical wall abut on bedding surfaces. In particular, almost all the Set 2 fractures abut on the bedding surfaces, identifying it a perfectly bed bound fracture set in the height classification scheme of Hooker et al. (2013). This implies that bedding surfaces, in this context, actually have a control on the

vertical development of the fractures. This is particularly true for the "high order" bedding surfaces that limit the bed package, where almost 50% of Set 1 fractures abut.

Set 2							
	Test result	**Mean dip dir.**	**Mean dip**	**Fisher K**	**Kent K**	**Kent β**	**n. data**
Orientation	rejected	75.37	89.31	26.47	27.91	3.1	1933
	Set 1	**Set 3**	**S (Bedding)**				
FBI	21.41%	9.69%	97.91%				
	Dist. Type	**Mean**	**Standard dev.**	**n. data**			
Length distribution	Lognormal	2.094	2.2	913			
Height distribution	Lognormal	1.3	0.78	32			
H/L ratio	0.359						
	Value	REA					
P_{21}	\	\					

Table 6 Result summary for Set 2 fractures.

Set 3 shows the highest number of abutting relationships compared to the other sets and the shortest average length, consistent with Set 3 being the most recent set in this outcrop.

Due to the unfavourable orientation of the vertical wall (sub-parallel to Set 3 average attitude), only orientation data from facets can be collected and the height distribution cannot be characterized. This implies that also the H/L ratio cannot be obtained. Finally, the same limitations as for Set 2 apply to P_{21} and REA calculations. Therefore, if this

fracture set were to be modelled with a stochastic approach, many relevant assumptions must be introduced even if, in general, the quality of the outcrop is high.

Set 3							
	Test result	**Mean dip dir.**	**Mean dip**	**Fisher K**	**Kent K**	**Kent β**	**n. data**
Orientation (Set 3a)	rejected	157.46	78.75	132.83	133.52	4.45	1269
Orientation (Set 3b)	rejected	124.08	76.8	91.83	92.63	4.2	2123
	Set 1	**Set 3**	**S (Bedding)**				
FBI	31.2%	21.19%	\				
	Dist. Type	**Mean**	**Standard dev.**	**n. data**			
Length distribution	Lognormal	0.74	0.75	1863			
Height distribution	\	\	\	\			
H/L ratio	\						
	Value	**REA**					
P_{21}	\	\					

Table 7 Result summary for Set 3 fractures.

11 Discussion

This contribution is focused on the geometrical and topological characterization of fracture networks and in particular on the input parameters necessary to generate stochastic DFN models. The main goal of this paper is to provide quantitative methodologies that minimize user choices through the implementation of statistical tests (e.g. orientation distribution). If statistical tests are not applicable due to the non-compliance with the underlying assumptions, other statistical parameters (e.g., P_{21} REA) or statistical distances from a non-parametrical estimator (length and height distribution) are proposed. The presented methodologies rely on data collected from DOMs, both point clouds and orthomosaics. In the context of upscaling geometrical parameters, DOMs are a suitable support for collecting data on large outcrops, reducing the time for the acquisition process, allowing data collection in areas that are inaccessible for practical or safety reasons, and opening up the possibility of implementing automatic feature extraction methods or automatic classification methods (topology). For a complete characterization of the fracture network, especially when performed for fluid flow simulations, the geometrical parameters included in this contribution must be integrated with more in-depth analysis, to characterize filling, mineralization and other characteristics of the network (e.g., microscale connectivity) that can be assessed with other type of techniques and at a smaller scale (Forstner et al., 2025). Our approach is based on a combination of field and DOM surveys, the latter used to obtain large datasets to support statistical analysis, the former to guide the digital mapping, filtering the noise given by external factors (e.g., quarrying operations) and assigning every fracture to a specific set through geological observations (particularly from kinematics and relative chronology).

In quarries, fractures develop during excavation. It is thus of the utmost importance, both for genetic reconstructions and analogue modelling, to exclude fractures that are related to anthropogenic surface processes. On our case study, the measured fracture sets are in accordance with the existing literature on the area (Sec. 2). In the outcrop pavement there are areas covered by debris accumulations, where no fracture set are detectable. Other parts of the pavement are affected by quarrying activities, resulting in zones “saturated” by fractures with randomly oriented and radially distributed fractures (related to balsting). Whitin these areas, only Set 1 is clearly detectable, given its constant spacing and orientation, its average length higher than the other fractures and its centimetrical-scale displacement. On the contrary, Set 2 and Set 3 are masked by the artificial fractures and even if they are present, it is difficult to reliably isolate and digitize them.

The high quality of the outcrop, with adjacent horizontal and vertical surfaces, was instrumental in testing techniques that represent, in our opinion, a step forward in collecting rich quantitative datasets and developing rigorous statistical treatments for many geometrical and topological parameters of a fracture network (Table 1). However, it should be noted that for some parameters, data collection and analysis still have limitations. These two points are addressed in the following subsections.

11.1 Combined analysis of fracture traces and faces

The integration of facets and traces (collected both on horizontal and vertical outcrops) allows a complete characterization of the parameters listed in Table 1, while other approaches rely on the analysis of facets or traces only (e.g. Ortega et al., 2006; Boro et al., 2014; Martinelli et al., 2020; Smeraglia et al., 2021).

Orientation data have been collected on the vertical wall PC-DOM, where dip and dip direction of true 3D planes can be measured by fitting a mean plane to planar patches of the point cloud.

TS-DOMs enable the digitalization of fracture traces and interpretation boundaries on both horizontal and vertical outcrops at a fixed scale, corresponding to the resolution at which they were collected. Nonetheless, the proposed methodologies can be applied to different scales, from thin sections to satellite images, provided that data are organized as digitized fracture traces combined with the interpretation boundary. Integrating the interpretation boundary with the fracture traces is fundamental to avoid underestimating the connectivity index by misinterpreting B-nodes as I-nodes and provides a fundamental input to identify censored fractures. P_{21} is calculated on the pavement TS-DOM where the huge areal extension (≈ 18.000 m2) enables to define a sufficient number of scan areas to detect the REA lower threshold.

11.2 Orientation analysis

The methodologies we propose for orientation analysis aim at reducing subjective choices of the interpreter while exploiting semi-automatic data collection to increase the volume of data that support statistical analysis. This was achieved by:

- Introducing cluster analysis, in addition to classical structural observation, to segment fracture sets following a statistical criterion.
- Calibrating an automatic feature extraction algorithm (FACETS) to maximize the data that can be extracted from PC-DOMs, avoiding the generation of artifacts (e.g. planes with an intermediate orientation), as sometimes occurs in workflows tested by previous authors (e.g. Menegoni et al., 2019; Panara et al., 2024).
- Rather than assuming circular symmetry and fitting a Fisher distribution without prior statistical verification, our approach explicitly tests the fitted orientation distributions using goodness-of-fit tests. This provides a statistically sound assessment of fracture set orientation parameters.

Our rigorous analysis revealed that, in our case study, no fracture set follows a Fisher distribution (Tab.5, 6, 7), in contrast to the conclusions of previous studies on the same outcrop (Panza et al., 2016; Zambrano et al., 2016). Moreover, Kent distribution parameters indicate a low ovalness for all fracture sets, even if Set 3a and Set 3b seem to be almost elliptical clusters (Fig 5e). The reason is that sets 2, 3a, and 3b are strongly asymmetrical and set 1 is multimodal, as highlighted by contour plots in figure 5.

The applicability, and thus the quality of the results produced by the automatic feature extraction algorithm, depend heavily on the ability to distinguish and characterize each fracture set within the network. In this study, reliable results were obtained by clearly distinguishing fracture sets through the integration of field data, DOM-derived data, and clustering analysis. In more geologically complex contexts, where fracture sets are less well defined, caution is advised both when applying the clustering algorithm—since the number of sets must be specified *a priori*—and when using the automatic feature extraction algorithm.

In the context of generating analogue fracture sets with stochastic modelling (e.g. in DFN), where only the Fisher distribution can be modelled by standard software, these parameters should be handled with care. Using a Fisher distribution with a K parameter small enough to fit all the planes in the cluster (i.e. with a large spherical variance) will result in artifacts along the minimum axis of the elliptical distribution, or in not properly represented tails in case of asymmetrical distributions. On the contrary, using a K parameter that is too large (i.e. with a small spherical variance) will result in an underrepresentation of the oval tails of the data. All these problems will result in an incorrect modelling of connectivity, that is positively correlated to orientation dispersion (e.g., Smith et al., 2013).

A possible solution to this problem, using available software, is to split the oval clusters and fit multiple Fisher distributions to model their parts, possibly validating the results generating synthetic clusters and comparing them with the natural ones. More generally, more advanced distributions such as the, Kent, Bingham-5-parameters, Bingham-8-parameters or mixed Bingham can be adopted to fit asymmetrical or multimodal clusters (Kurz et al., 2014; Gilitschenski et al., 2016; Yamaji, 2016), in our experience adopting at least the Kent distribution in stochastic modelling applications seems to be a significant improvement toward creating realistic stochastic fracture networks.

11.3 Topology

In the ongoing discussion on topological analysis in fracture studies, some authors (e.g. Sanderson and Nixon, 2015) propose considering the connectivity of fracture branches, instead of full fracture traces, due to a expected uncertainty in unravelling crosscutting, abutting, or splay relationships when branches form a small angle. Other authors (Forstner and Laubach, 2022) suggest considering also contingent nodes (C nodes) that would allow individual small branches to merge into larger traces, depending on the considered scale and/or diagenetic consideration. In this context, it is worth recalling that (i) linear traces or branches represent the intersection of fracture surfaces with the outcrop surface (e.g. Sanderson and Nixon, 2015), and that (ii) often, it is the fracture surfaces themselves that are of interest, not their traces. In other words, branch connectivity parameters and length distribution can provide useful information to characterize 2D connectivity of lines, but it is important to enphasize that considering branch length data as if they were full trace lengths would dramatically underestimate the dimensions of the underlying fracture surfaces, overestimate fracture density (since branches are more numerous that full traces) and would not allow performing directional topology analysis by fracture set, as discussed above.

Therefore, we emphasize the importance of considering fracture traces, relying on geological information collected in the field to solve the ambiguity highlighted by Sanderson and Nixon (2015).

Directional topology improves the fracture network characterization by assigning every node to one or more specific fracture set(s). This allows for the quantification of crosscutting and abutting relationships between different fracture sets and an understanding of how the different fracture sets contribute to the overall connectivity of the fracture network. Simultaneously, it is possible to derive parameters like the FBI that express the relative chronology and relationships with bedding surfaces more quantitatively. The necessity of excluding censored fracture traces from the FBI calculation might result in either an overestimation of this index, in case the longest traces tend to have I-nodes, or an underestimation in case they tend to terminate with Y-nodes. Unfortunately, this effect is difficult to assess at the moment, and a more detailed analysis will be proposed for future studies.

Extracting the backbone of the trace network as shown in Figure 9, i.e. the largest connected cluster of the network, appears to be an interesting result, made possible by the code we developed (https://github.com/gecos-lab/FracAbility). However, we would like to raise a few cautions about these results. Indeed, in the studied outcrop the backbone is located along one of the major structures present in the pavement but, at the same time, it was detected in a particularly clean area of the pavement, furthermore, free from quarrying-induced fractures. This area is bounded by zones where the detection of smaller fractures belonging to Set 2 and 3 is more limited. Therefore, it is not possible to determine if the shape and position of the backbone are only due to the presence of a major structure, or whether they would be different if other areas of the outcrop were undisturbed. In conclusion, the backbone detected by our analysis probably represents a subset of what it could be, if a complete undisturbed dataset was available. This also means that the connectivity index of the entire fracture network is probably underestimated, even in such a high-quality outcrop.

11.4 Length and height distributions

Of the four major biases that hinder the definition of a correct length and height distribution, censoring is the only one that cannot be addressed by changing sampling strategy (as for size and orientation bias) or by improving data acquisition techniques and the quality control of input data by quantifying the resolution of the TS-DOM relative to the smallest detectable fracture (truncation bias, Section 7). This has led to the exclusion of censored data (Bisdom et al., 2014) or to consider censored fractures as complete ones (Panara et al., 2024; Smeraglia et al., 2021), resulting in statistical models that always underestimate lengths (Benedetti et al., 2025). Furthermore, non-parametric approaches do not provide a fully specified parametric distribution (Mauldon et al., 2001), that is a fundamental input data in applications such as stochastic models, and also for evaluating the meaning of data in general (i.e. knowing the mean without any hint on standard deviation is meaningless). The censoring correction obtained by

Benedetti et al., (2025) with survival analysis makes it possible to use the full extent of the dataset collected from TS-DOM, and to obtain an unbiased statistical distribution.

In this contribution, we leverage the availability of both pavement and vertical wall exposures, obtaining an H/L ratio specific for our outcrop. This is a significant improvement with respect to previous approaches, where the H/L ratio is generally assumed based e.g. on theoretical mechanical considerations, without any comparison with empirical data (Odling et al., 1999; Schultz and Fossen, 2002, and references therein).

The approach we propose, however, is not entirely free from *a priori* assumptions. Indeed, even in ideal outcrops, it is impossible to actually measure H/L of a single fracture surface because either the direct connection of the fracture surface in the vertical and pavement exposures is observed – but in this case, both traces are censored, or complete traces are observed – but in this case the connection is invisible. We assume that height and length statistics are correlated according to size rank, as previously stated in Section 9. It seems to us that this is a reasonable assumption, because it may be based on mechanical considerations (Odling et al., 1999) and in the light of results of the linear regression proposed in Figure 15. However, it is nevertheless important to state this transparently and to keep in mind that when using the resulting estimate in fracture modelling.

11.5 Fracture intensity and representative elementary area

Areal fracture intensity is often estimated using methods based on scan lines, scan areas or circular scan line (Rohrbaugh Jr. et al., 2002; Zeeb et al., 2013). These methods provide a minimum scan area size for a representative estimation of P_{21} based on the mean fracture trace length. In this contribution, we propose a different approach, based on the concept of Representative Elementary Area (REA) to try to quantify the range of scan area size for which fracture intensity can be mediated to ensure a proper continuum-equivalent description (Bear, 1975).

As a partial correction to the approach in Martinelli et al. (2020), we recently noticed that the finite nature of outcrops determines limits in the collection of P_{21} data, such as the progressive decrease in number of scan areas as their size increases, and the non-independence of P_{21} samples collected from scan area of increasing sizes. For these reasons, a quantitative approach based on formal statistical tests cannot be reliably applied. Adopting a more qualitative approach will result in a less significant result, which partially depends on the interpreter choice. However, this approach does not require assumptions as restrictive as those used by Martinelli et al., (2020). We also recognize that adopting a more qualitative approach may introduce subjectivity in the selection of window size, but still having an order of magnitude for the REA (and hence for REV) is important in modelling studies. Indeed, in addition to formal statistical reasons, defining the REA has important practical applications. For instance, the choice of the optimal cell size in reservoir-scale models stems from a combination of several factors, that are the geological characteristics of the area, lithostratigraphic heterogeneities, mean spacing between wells and the

available computational power. In general, however, there is a lower limit to the resolution of the model - around 50 m of cell size, dictated by computational power. This is an important piece of information because it outlines the minimum size that an analogue outcrop must have to capture all the variability within a hypothetical cell. From this point of view, the Pontrelli quarry provides a pavement two to three times larger than the minimum cell size, granting a sufficient area to calculate representative statistics. At the same time, the P_{21} REA, determined for Set 1 fractures between 5m and 12m, is approximately five to ten times smaller than the minimum cell size, allowing a safe application of continuum-equivalent upscaling techniques.

12 Conclusions

In conclusion, this paper presented a series of quantitative methodologies to characterize fracture network geometry and topology from Digital Outcrop Models (DOMs). The ideal conditions for applying our methodologies involve an outcrop that enables the collection of a statistically significant and complete dataset (depending on the scope of the work). This requires favorable orientation of the outcrop faces relative to the fracture set orientation, overall surface cleanliness (minimal debris, vegetation, or damaged zones), sufficient size to ensure adequate sampling, and the presence of at least two perpendicular exposures (horizontal and vertical). Although such conditions are challenging to achieve in natural settings, they should serve as guidelines for selecting a suitable outcrop. Among all the parameters required to fully characterize a fracture network, we focused on those required to generate 3D stochastic DFN models, that are: orientation parameters, topological relationships, length and height distribution parameters, H/L ratio and P_{21}:

- Orientation data are collected through a semi-automatic workflow, clustered with k-medoids, and tested for the goodness-of-fit to a Fisher distribution. Alternatively, the Kent distribution parameters are also provided. This procedure allows subjectivity to be removed from the assignment of dip/dip direction data to a specific fracture set and supports the choice of meaningful orientation parameters through the implementation of statistical tests.
- Topological relationships are calculated including the interpretation boundary, this allows to: (i) to define B nodes and exclude them from the connectivity index (CI) calculation (ii) to identify censored fractures in an automatic way. Backbone extraction highlights the presence of large, connected clusters in the network. Crosscutting and abutting relationships between different fracture sets are quantified through directional topology.
- The approach developed to deal with censoring bias provides as a result a set of fully specified distributions (all parameters are explicit) corrected for censoring. The best model among the initial selection is defined through a graphical approach and a series of statistical distances.

- We demonstrate that estimating H/L is not possible without introducing some assumption, even for the best exposed set and in the presence of both horizontal and vertical exposures. Therefore, we opted to make our assumption as transparent as possible, and we tested it with regression analysis.
- P_{21} REA is calculated with a qualitative approach to avoid violating the underlying assumption of more formal statistical tests.

13 Code and data availability

Codes and data are available at the following GitHub repositories owned by the Gecos-lab group of the University of Milano – Bicocca (https://github.com/gecos-lab):

- FracAttitude: Python code for orientation data analysis available at https://github.com/gecos-lab/FracAttitude;
- DomStudioOrientation: Matlab code for orientation data analysis available at https://github.com/gecos-lab/DomStudioOrientation;
- FracAbility: Python toolbox for topology and survival analysis available at https://github.com/gecos-lab/FracAbility;
- FracAspect: Python code for H/L ratio calculation available at https://github.com/gecos-lab/FracAspect;
- FracElementary: Python code for P_{21} and REA analysis available at https://github.com/gecos-lab/FracElementary.

Chapter 2 - Unbiased statistical length analysis of linear features: Adapting survival analysis to geological applications

This chapter has been published in Solid Earth as https://doi.org/10.5194/se-16-367-2025

Abstract. A proper quantitative statistical characterization of fracture length or height is of paramount importance when analyzing outcrops of fractured rocks. Past literature suggested adopting a non-parametric approach, such as circular scanlines, for the unbiased estimation of the fracture length mean value. This is due to the fact that, in the past, estimating any type of statistical distribution was difficult and there was no real interest in defining precise parametrical models. However, due to the recent raise in popularity of Digital Outcrop Models (DOMs) and of stochastic Discrete Fracture Networks (DFNs), there is an increasing demand for distribution-based solutions that output a correct estimation of parameters for a given proposed model (e.g. mean and standard deviation). This change in demand highlights in geological literature the absence of properly structured theoretical works on this topic. Our methodology, presented for the first time in this contribution, represents a powerful alternative to non-parametrical methods, aimed at specifically treating censoring bias and obtain an unbiased trace length statistical model. As our first objective, we propose to tackle the censoring bias by applying survival analysis techniques: a branch of statistics focused on modelling time-to-event data and correctly estimating model parameters with data affected by censoring. As a second objective we propose a novel approach for selecting the most representative parametric model. We combine a direct visual approach and the calculation of four statistics to quantify how well proposed models reflect the data. We apply survival analysis to correctly estimate statistical parameters of censored length dataset in three different case studies and show the effects of censoring percentage on parametrical estimations that do not use this paradigm. The presented analyses are carried out using the open-source Python package called FracAbility, that we purposefully created to carry out the described workflow (https://github.com/gecos-lab/FracAbility).

1 Introduction

Fractured rock masses are complex systems composed of intact rock and discontinuities (Hoek, 1983). Characterizing the statistical distribution of 3D geometrical properties of such discontinuities (e.g. aperture, roughness, area, orientation, height/length ratio, etc.) is fundamental for understanding and modelling mechanical and hydraulic properties of rock mass and fluid-rock interaction. The recent increase in computing power and the emergence of new approaches based on Digital Outcrop Models (DOMs) allow for the extraction of large datasets

and facilitate the sampling of properties instead of just their estimation (Bistacchi et al., 2015; Tavani et al., 2016; Healy et al., 2017; Thiele et al., 2017; Marrett et al., 2018; Nyberg et al., 2018; Bistacchi et al., 2020; Martinelli et al., 2020; Bistacchi et al., 2022a; Mittempergher and Bistacchi, 2022; Storti et al., 2022). The need for a more rigorous statistical approach to structural data analysis is also motivated by the popularization of stochastic Discrete Fracture Networks (DFNs) as a modelling approach for rock masses. In DFNs, discontinuities are represented, in a simplified way, as finite planar surfaces, generally rectangular, polygonal or elliptical (Cacas et al., 1990; Dershowitz and Herda, 1992; Tavakkoli et al., 2009; Hyman et al., 2015). In fact, stochastic algorithms used to generate these surfaces are guided by statistical distributions obtained from field or well data, or assumed based on some prior knowledge (Andersson et al., 1984; Cacas et al., 1990; Davy et al., 2018). For example, DFNs where fractures are represented as rectangular surfaces (the most common implementation) require defining fracture size via parametric distributions of length (measured along strike) and height (measured along dip), or alternatively length and length/height ratio (Hyman et al., 2015). However, it is worth noting that DFNs are not the only viable approach to model fractured rock volumes. Other methods such as tensorial approximations (Brown and Bruhn, 1998; Suzuki et al., 1998) based on the crack tensor measure (Oda, 1983) are also used (Healy et al., 2017) and, with details depending on the implementation, require similar parameters in input. The main obstacle that geologists encounter while trying to characterize a fractured rock volume is the impossibility of directly measuring the 3D properties of discontinuities, since only indirect geophysical methods may provide complete 3D datasets. However, imaging discontinuities with geophysical methods is strongly limited by spatial resolution and/or by the absence of contrast in the physical properties investigated by a particular technique (Martinelli et al., 2020). Therefore, a vast body of research focuses on the characterization of properties of discontinuity traces or lineaments, i.e. the 2D intersections of 3D discontinuity surfaces with the outcrop surface, or with topography (Dershowitz and Herda, 1992; Bonnet et al., 2001; Manzocchi, 2002; Bistacchi et al., 2011; Sanderson and Nixon, 2015; Bistacchi et al., 2020; Martinelli et al., 2020; Storti et al., 2022). In this contribution, we focus on the problem of defining accurate and unbiased length (or height) distributions based on data collected on DOMs. In any given outcrop, lineament length measurements will always be affected by four main biases: length (i.e. size), censoring, truncation and orientation (Baecher, 1983). The correction of these biases has been thoroughly researched, and the standard solution currently adopted by many authors is based on circular scanlines (Mauldon, 1998; Zhang and Einstein, 1998; Mauldon et al., 2001; Rohrbaugh Jr. et al., 2002). The method consists of drawing a circle directly on the outcrop or on an image and counting the intersections of the circle with lineaments. With a chain of assumptions, defining an indirect relationship between mean length and the intersections of lineaments with the scanline, an unbiased non-parametric estimation of mean length, fracture density and intensity can be obtained (Mauldon et al., 2001). Thanks to its simple implementation in the field, this technique is widely used, however it has an important limitation: lineament lengths are never directly measured. For this reason, analysis carried out with the circular scanline method yield estimates of mean length values, but are unable to provide a complete characterization of the

lineament length distribution and therefore lack any real statistical significance. Despite this problem, the popularity of this method was motivated by the fact that in the past, without modern digital imaging techniques, length data acquisition was slow and tedious and thus datasets were usually small. Moreover, calculating the length and estimating any distribution other than the exponential, was difficult and computationally intensive (Baecher and Lanney, 1978; Baecher, 1980) and, due to limitations in early algorithms used to generate stochastic fracture networks, there was no real interest in using precise distribution parameters in input. These limitations have been mostly overcome by modern characterization methods and thus new tools and techniques based on Maximum Likelihood Estimation such as FracPaQ (Healy et al., 2017; Rizzo et al., 2017) have been developed, enabling researchers to apply quantitative statistical inference on dense digital datasets. Our methodology, presented for the first time in this contribution, represents a powerful alternative to circular scanline methods to treat specifically the censoring bias and obtain unbiased trace length statistical models. This specific bias is defined when, for some traces, one or both ends cannot be seen due to the limited size of the outcrop (Baecher and Lanney, 1978). This effect is present from thin section to satellite image scale, and it is caused by the inability to see beyond the study area (i.e. thin-section limits, outcrop extension and so on; Mauldon et al., 2001). From a statistical point of view, this problem is analogous to the censoring bias affecting some medical, biological, and engineering datasets, and the techniques used in these disciplines to solve or limit the effects of this bias go under the names of survival analysis, life testing, or reliability analysis (Kaplan and Meier, 1958; Leung et al., 1997; Lawless, 2003; Cox, 2017; Karim and Islam, 2019). Even though in these disciplines the recorded random variables are time spans (e.g. lifetime of a patient, time-to-failure of a mechanical part, etc.), we demonstrate that this statistical technique can be also adapted to length measurements. Measuring length is straightforward with dedicated code or with a simple GIS software however, applying survival analysis and fitting robust unbiased parametric statistical distributions, needs a more detailed treatment. As the main topic of this contribution, we propose to adapt survival/reliability analysis techniques to correctly account for censored lengths and estimate robust trace length distributions derived from DOMs. Furthermore, we define a quantitative methodology to select the most representative estimated statistical model (i.e. parametric distribution) from a list of proposed models. The theory and techniques presented in this paper are available as an open-source Python package called FracAbility that accepts shapefiles as input and allows to carry out a complete and unbiased statistical analysis workflow for fracture length data (https://github.com/gecos-lab/FracAbility).

2 Fracture surveys and terminology

A discontinuity in a rock mass can be defined as a surface across which a material has lost its cohesion or was originally discontinuous. This definition thus includes faults, fractures, foliations, stylolites, compaction or deformation bands, and bedding interfaces. Fractures are classified as shear or tensional fractures. Tensional fractures, when possible, can be further divided into joints when empty or veins when filled by minerals (Davis et

al., 2011). This distinction however is often difficult to make without an in-depth field and sample validation, since some fractures may have hybrid fill attributes and may be only partly filled with inconspicuous mineral deposits and thus resemble joints. Moreover, the degree of fill may depend on fracture width, so that small fractures resemble veins (Laubach et al., 2019). Although three dimensional by nature, most of the times discontinuities are mapped as 2D lineaments or traces over a surface. These lineaments are the intersections of such discontinuities with an exposed surface, such as the topography, an outcrop, a borehole or a sample. In this work, following a common usage in outcrop studies, the term fracture is used as a generic term to indicate any type of discontinuity trace. Fractures with the same formation age, kinematics and orientation, can be grouped into families or sets. Multiple fracture sets present within an area form a fracture network or system (Davis et al., 2011).

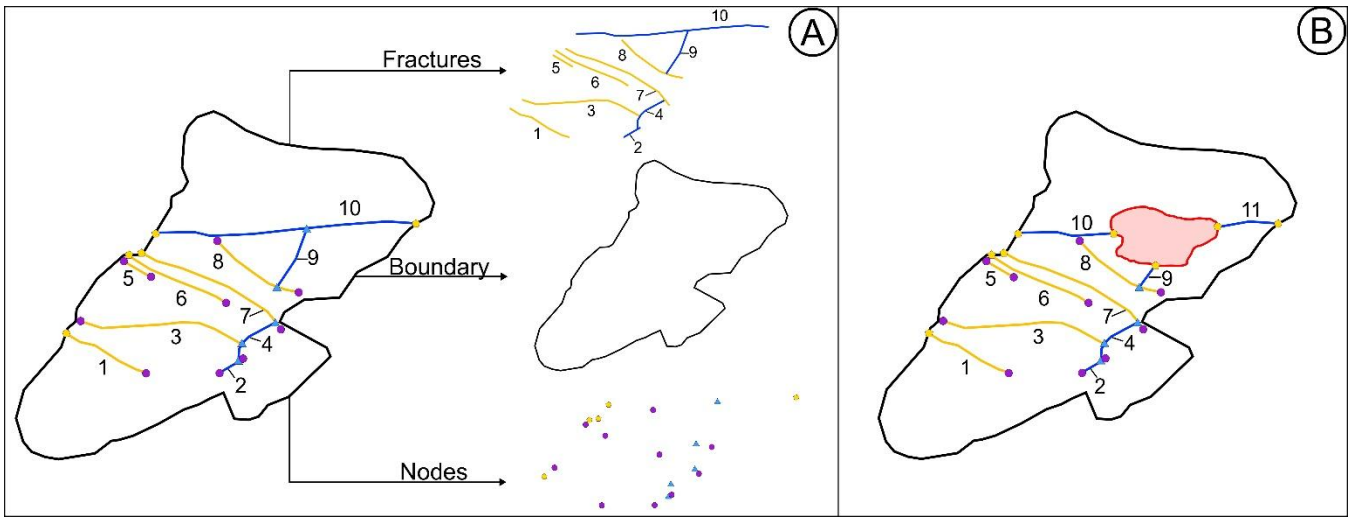


Figure 1 In (A) an example of a simple fracture network and its components. In (B), a modified version of the boundary with a "hole" (in red) in which no interpretation can be carried out. The presence of these holes can increase the number of censored fractures and introduce uncertainty on the interpretation, splitting fractures in multiple pieces (i.e. fracture 10 is split in two).

Additionally, fracture networks present two other fundamental components: Boundaries and Nodes (Fig. 1A). Boundaries are the limits of the sampling area, at scales from thin sections to satellite images, within which the sampled fracture traces are assumed to be complete. Ideally, boundaries are strictly convex, however this is often not the case. Boundaries often show tight bends, coves and holes and the final shapes are mostly controlled by localized alteration, anthropogenic activity, vegetation, etc. (Fig. 1B).

Nodes are points in the network that define how fractures interact or do not interact with other elements of the network (other fractures, boundary, holes). Nodes can be classified by the number and type of segments (branches) that insist on the given point (Sanderson and Nixon, 2015, Fig. 2):

- Isolated I nodes are connected to only one branch belonging to a fracture.
- Y/T and X nodes are respectively connected to 3 and 4 branches belonging to two fractures.
- "Boundary" B nodes are connected to 1 branch belonging to a fracture and 2 branches belonging to the boundary.

Node elements describe useful hydraulic properties of a fractured rock mass (Gueguen et al., 1991; Manzocchi, 2002; Sanderson and Nixon, 2015). In a non-porous rock with all open fractures, a network with a prevalence of I nodes is less connected and thus fluid flow is usually more restrained. Conversely, both Y/T and X nodes increase the connectivity of the network and thus increase both fluid flow and the permeability. However, sometimes this is indeed not the case e.g. sealed faults and opening-mode fractures (Forstner and Laubach, 2022). Moreover, if a rock is porous then length becomes the key parameter for controlling fluid flow and thus even in a network with only I nodes length can markedly augment fluid flow (Philip et al., 2005). Additionally, nodes give important chronological information about the network.

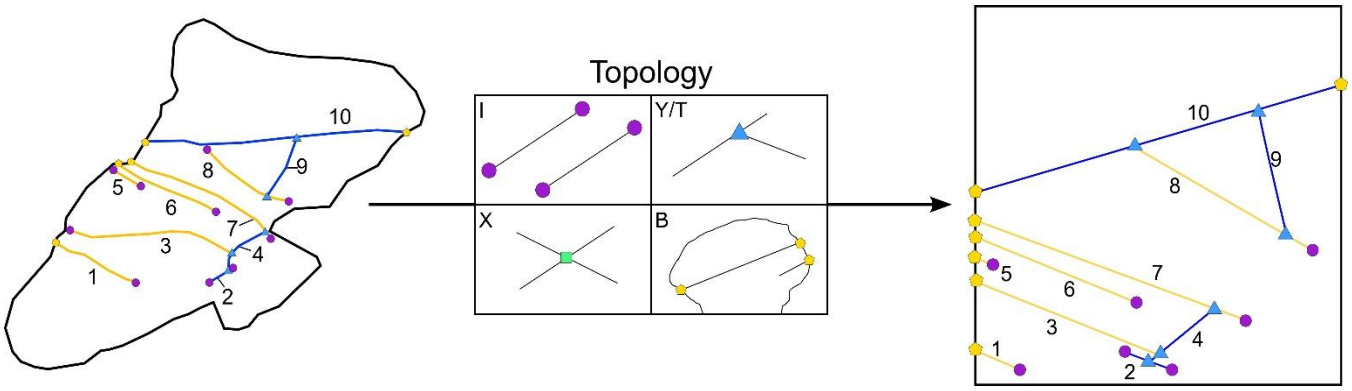


Figure 2 Topological abstraction of the simple fracture network in Fig. 1 following the described node classification.

I nodes indicate that fractures terminate within the observational boundary and that they do not end nor cross with other fractures. Y/T nodes can be an evidence of to either abutting or crosscutting relationships of non-coeval discontinuities; the former relevant for discontinuities with no displacement while the latter applies to discontinuities with visible displacement. X nodes represent a mutual crosscutting relationships and the chronological interpretation depends also on other observations. Finally, B nodes identify whether a fracture is complete or censored and as such define if measured properties along a trace are also complete or incomplete. These nodes could be considered as a sort of equivalent of the U (undefined) nodes in Nyberg et al. (2018) however with a slightly different meaning and more precise definition.

3 Statistical modelling of censored length data

Having a length dataset, there is usually the necessity of estimating the parameters of one or multiple statistical distributions. When doing so, censoring is inevitable as the area within which measures are carried out will always be limited. Then, how can we carry out an unbiased estimation of such parameters? On the one hand, common and simplistic approaches such as ignoring censoring (i.e. considering censored lengths as if they were complete) or excluding censored measurements (i.e. cherry-picking only non-censored data) should be avoided as they will always lead to an underestimation of the model parameters (see discussion for a more in-depth analysis). On the other hand, circular scanline methods offer an unbiased estimate of the mean length, however, being non-parametric, they do not yield neither the distribution type (e.g. normal, exponential, etc.), nor distribution shape parameters (e.g. standard deviation, variance, etc.). This in turn makes the estimate's use quite limited and not apt to downstream statistical modelling applications (such as DFNs). To solve these problems, we propose to use survival analysis, a specialized field of statistics, developed to deal with censored data. Survival analysis focuses on the analysis of time of occurrence until an event of interest (Kalbfleisch and Prentice, 2002). Although in literature the terms survival times, time-to-failure, or more generally lifetimes (Lawless, 2003) seem to imply that time is the only valid variable, any non-negative continuous variable, such as length, can be valid (Kalbfleisch and Prentice, 2002; Lawless, 2003). The advantage of survival analysis over the other methods discussed above is that it considers censored data as a valid datapoint, carrier of the information that the event did not occur up to the censoring time.

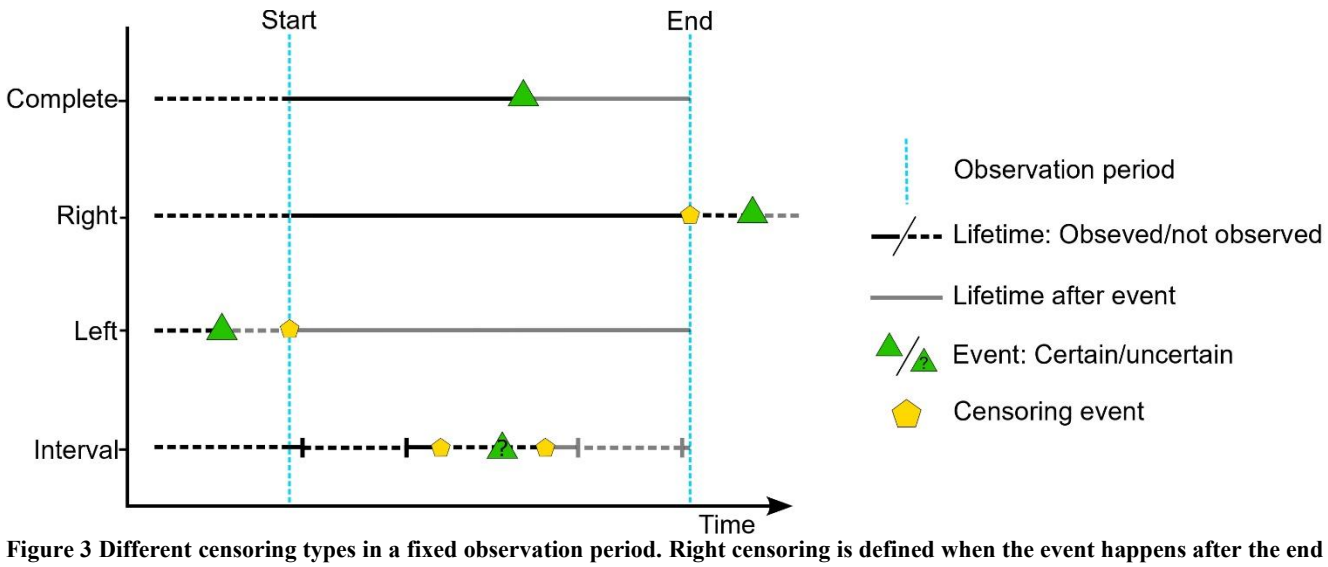


Figure 3 Different censoring types in a fixed observation period. Right censoring is defined when the event happens after the end of the study. Left censoring occurs when the event happens before the start of the study. Interval censoring happens somewhere between observation intervals within the study period.

This is a necessary shift in perspective that allows for an unbiased estimation of statistical models that will be described in the following. We will start by describing the theory of survival analysis in function of time, and then we will show how the same theory can be applied in space, to sets of length or distance measurements.

3.1 Survival analysis theory and standard terminology

Since this technique is rooted in medical and biological applications, the nomenclature from this type of literature is carried along. The event of interest (for which we measure the time-to-event) is often defined as death, while a loss indicates that the observation has been lost because it was hindered by a secondary event, called a censoring event (Kaplan and Meier, 1958). Censoring events can be classified in three main types depending on when censoring happens in respect of the observation period (Karim and Islam, 2019, Fig. 3):

- Right censoring: The event happens after the end of the study period and thus the partial lifetime of the event is observed.
- Left censoring: The event happens before the start of the study and thus is not observed.
- Interval censoring: The event happens somewhere between observation intervals within the study period, because it was not possible to continuously monitor the occurrence of the event.

The most treated and common type is right censoring, that can be further divided in (Fig. 4):

- Single censoring: When there is an imposed end time equal for all events (i.e. controlled loss).
- Random censoring: When each measured event is characterized by a random censoring event (i.e. accidental loss).

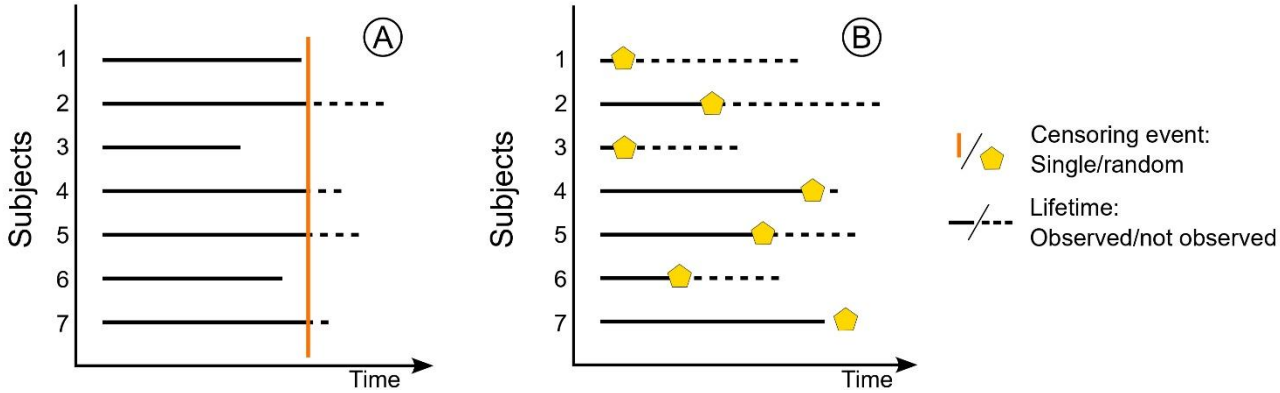


Figure 4 Different right censoring types for the same seven subject experiments. In (A) represented single censoring where events 1, 3 and 6 are complete measurements while the remaining are censored at time C. In (B) only event 7 is complete while the others are all censored at different times.

3.2 The survival curve and the Kaplan-Meier estimator

A fundamental assumption in survival analysis is the independence of the right censoring mechanisms, i.e. the assumption that the probability of occurrence of the event does not depend on the censoring mechanism (Kalbfleisch and Prentice, 2002; Lawless, 2003; Kleinbaum and Klein, 2012). Independent censoring is also called non-informative censoring as it does not affect inference and only indicates that the lifetime exceeds the censoring time (Kalbfleisch and Prentice, 2002). This assumption is the basis of the non-parametric Kaplan-Meier estimator of the survival function (SF), i.e. the population probability that X_i exceeds a given value x (Kaplan and Meier, 1958; Kalbfleisch and Prentice, 2002):

$$SF(x) = P(X_i > x) \quad [1]$$

The survival function is often called complementary cumulative function, since it adds to 1 with the cumulative distribution function (CDF):

$$CDF(x) = P(X_i < x) = 1 - P(X_i > x) = 1 - SF(x) \quad [2]$$

The non-parametric Kaplan-Meier estimator $\hat{P}(x)$ is defined by ordering N values of complete x_{co} and censored x_{ce} data by increasing magnitude, with $x_1 \leq x_2 \leq \cdots \leq x_N$, such that (Kaplan and Meier, 1958; Kalbfleisch and Prentice, 2002):

$$\hat{P}(x) = \prod_r \frac{N-r}{N-r+1} \quad [3]$$

where r are the indexes in which the complete values are smaller than a value x (i.e. $x_{co} \leq x$). Therefore, $\hat{P}(x)$ is a step-function that (Fig. 5):

- remains constant in any given time interval where no new events are recorded or where censoring occurs.
- decreases, by a degree (step) depending on the number of events that occur in a time interval.

Empirical survival curves are fundamental for representing, comparing, and understanding the survival rates of different censored datasets. The curve's steepness is directly proportional to the survival rate and the median of the (random) survival time (i.e. time value corresponding to a 50% chance of survival) can be used as a simple indicator to compare different survival curves.

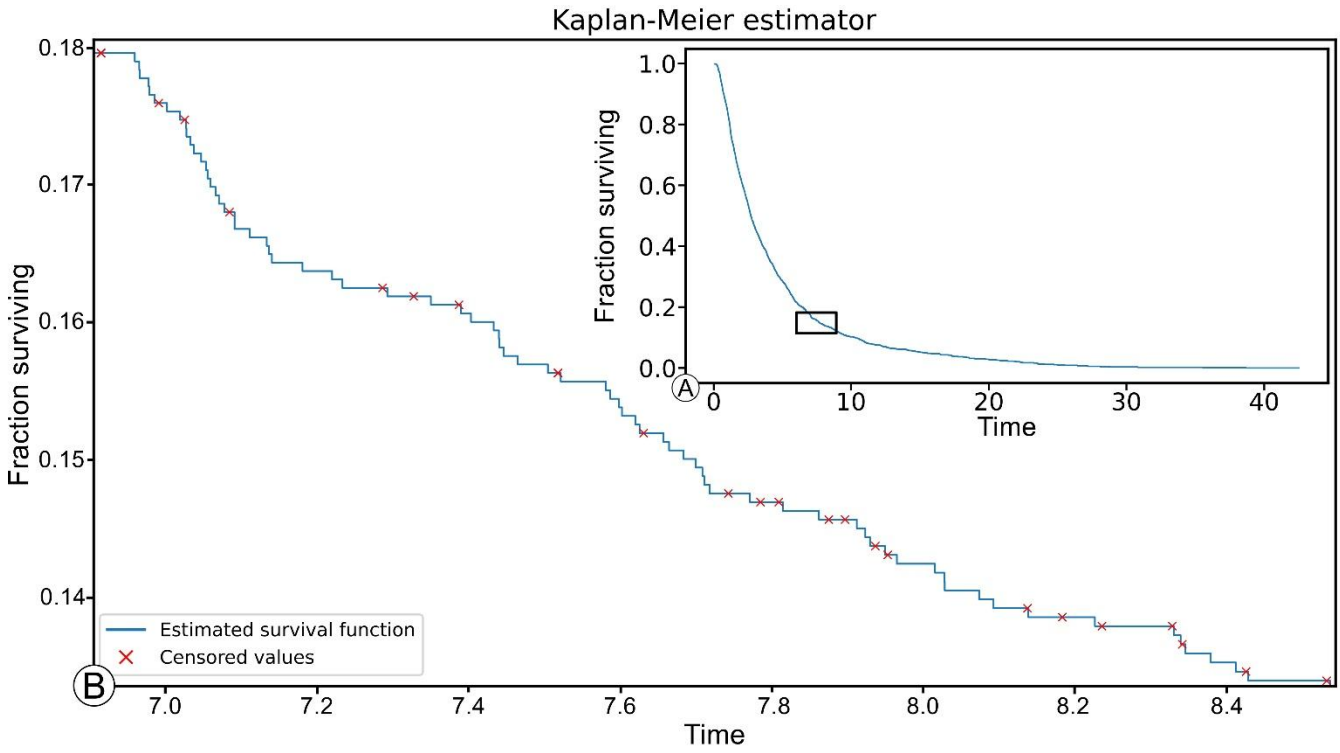


Figure 5 An example of the Kaplan-Meier estimated survival curve. From A the estimated curve seems continuous however by zooming in (red rectangle in A) it is a step function (B). The red X indicate the censoring events and show how the function never jumps at these values.

3.3 The time-length dimensional shift

Since length is, as time, a non-negative continuous variable, it is theoretically possible to apply survival analysis techniques to length datasets by considering (Fig. 6):

- the complete fracture trace length analogous to the lifetime.
- the event as the end (death) of a trace, marked by a node (I and/or Y/T nodes in case of complete traces).
- the study area, defined by its boundary, analogous to the study period.
- the censored event as the intersection between the fracture trace and the boundary (marked by a B node).

By applying the definitions of the different types of censoring (described in 3.1) to our specific application, it is reasonable to assume that only random censoring occurs in trace length analysis. Moreover, censoring is non-informative since the boundary is the product of secondary events (i.e. alteration, erosion, debris covering parts of the outcrop, vegetation, human activity, etc.) that occur after fracture genesis and thus do not inform the occurrence of the event (see the discussion for a more in-depth analysis).

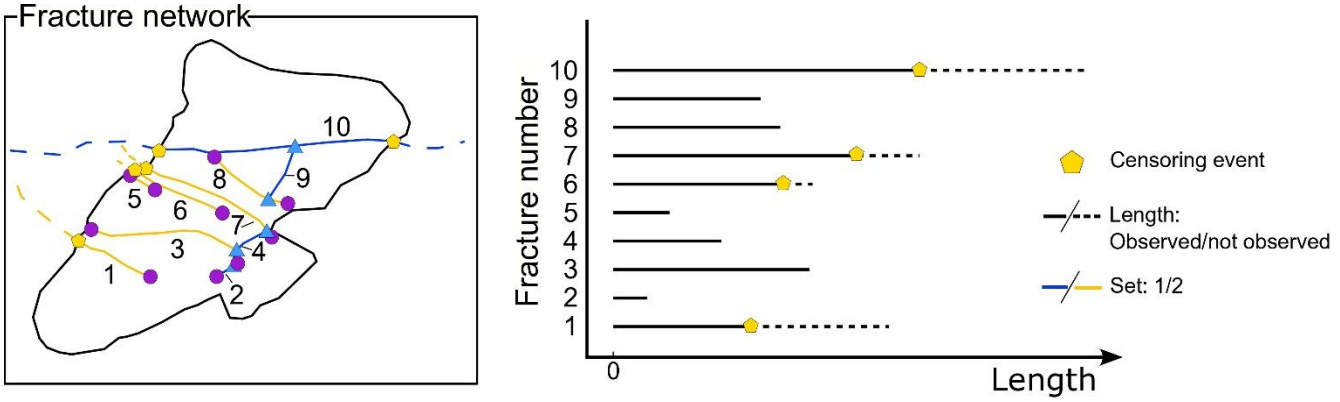


Figure 6. Censoring effect on an example of a simple fracture network and corresponding survival diagram. The survival diagram is a 1D representation of the fracture length. On the Y axis the fracture number is indicated and on the X axis the length is represented. Solid lines indicate the actual measured length while dashed lines indicate the possible continuation of the fracture. Yellow pentagons represent the censoring event given by the intersection of the fracture with the boundary.

3.4 Unbiased MLE model estimation for censored data

Considering that the censored event does not modify the probability of occurrence of the measured event, survival analysts try to solve the problem of estimating a parametric statistical model (i.e. distribution) by using optimization algorithms such as linear regression or the Maximum Likelihood Estimator (MLE). Both methods are valid, however in these types of applications MLE is most used and thus we will discuss more in detail this technique. Given a statistical parametric model with density $g(x, \theta)$ (i.e. an assumed theoretical distribution) and a sample x of size n, the main objective of MLE is to estimate the parameters $\hat{\theta}$ such that the observations x are the most likely under $g(x, \hat{\theta})$ (Anderson and Burnham, 2002; Burnham and Anderson, 2004; Karim and Islam, 2019). In the simplest 1D case (estimation of the parameter of a one-parameter distribution), the likelihood can be described as a chain product of probabilities carried out on all n individuals in the sample:

$$L(\theta|x, g) = \prod_{i=1}^{n} g(x, \hat{\theta}) \quad [4]$$

Given the fixed (assumed) distribution type g with parameter θ and a sample x, the likelihood will be the product of all the probabilities $g(x, \hat{\theta})$, i.e. the distributions Probability Density Function (PDF) calculated at value x_i $\left(P\ (\theta|x_i)\right)$. In practice, since g and x are fixed, the likelihood is function only of the distribution parameter θ. The maximum of this function will be $\hat{\theta}$, i.e. the parameter that maximizes the likelihood for that model type and data combination. This process can be extended to estimate multiple parameters, increasing the complexity, but the core concept remains unaltered. However, Eq. (4) is valid only for complete datasets because the probabilities associated with right-censored data cannot be calculated using the PDF. To solve this problem, it is possible to

calculate the MLE for censored datasets, under the assumption of random censoring, using the models Survival Function (SF) $S(\theta|x_i)$ instead of the PDF for the censored data (Karim and Islam, 2019):

$$L(\theta|x,g) = \prod_{i=1}^{n} P(\theta|x_i)^{\delta_i} \times S(\theta|x_i)^{1-\delta_i} \quad [5]$$

Where δ is an on-off switch for complete ($\delta = 1$) or censored values ($\delta = 0$). In other words, when a datum is censored, we use the probability information that the individual survived up to the censoring event, given by the survival function. MLE is a consistent (i.e. estimation approaches the population parameter as sample size increases) and efficient (i.e. lowest variance) estimator (Enders, 2005), however it has its limitations. For example, if the model has more than one parameter, the weight of influence of each parameter is not known and thus it is difficult to know which parameter controls the fit. Moreover, simplistic MLE algorithms output a single likelihood value, with a possibility that the optimization gets trapped in a local minimum. This ultimately leads to questioning whether the estimated parameter or combination of parameters is the absolute best or if there are other more optimal solutions. These (and other) problems culminate to an even greater uncertainty related to the choice of the distribution, as an infinite variety of functions can be used for the same dataset. These uncertainties tie nicely to the following chapter discussing model selection comparison criteria and how we propose to solve this problem.

3.5 Model selection

With a series of fitted models, it is important to understand which model better represents the data. In geological literature this is usually archived using a specific type of null hypothesis tests defined as goodness-of-fit (GoF) test (i.e. Storti et al., 2022). These types of tests, in general, do not identify the best fitting distribution among a set of possible distributions. A GoF test takes a distribution L and the sample data, and tells if the model L is not plausible, that is, if the probability of drawing such a set of data, from a population of distribution L, is too small. This means that more than one distribution may be deemed plausible and the conclusion is strong only when a distribution is an unlikely model for the data. Moreover, GoF tests usually have underlying assumptions that if ignored undermine the test's accuracy (Storti et al., 2022). For example, the Kolmogorov-Smirnov test is biased if the parameters of the tested distribution are estimated from the data (Bistacchi et al., 2020). As an alternative to GoF tests, we propose a combined visual and quantitative approach that guides the researcher in an informed choice of the most representative model out of a list of sensible candidates. By sensible candidates it is intended those models that make physical sense for the case study. For example, for fracture lengths, a power law can be considered a sensible candidate since it describes an observed pattern of fracture self-similarity. A lognormal is also an equally sensible candidate because of the effect of truncation. Conversely a normal distribution is not very sensible since values in

the left tail can be negative. Thus, for a normal model, either the average length is very high and standard deviation very low or a truncated normal should be tested. In the case studies section, we will cover different models purposefully adding non-sensible models (such as a standard normal) to showcase the model selection workflow.

3.5.1 A visual approach using the Probability Integral Transform (PIT)

The Probability Integral Transform (PIT) is a well-known transformation of continuous distributions which converts random variables with continuous distribution to standard uniform distributed random variables (Fisher, 1990). Indeed, given:

- A random variable Y.
- F_Y the CDF of Y.
- $Z = F_Y(Y)$ is the PIT of Y and is a standard uniform random variable.

This means that, given a random sample X, sampled from a population with distribution Y, the transformed sample $Z = F_Y(X)$ has, with large probability, an empirical distribution which is close to the standard uniform distribution. Clearly, if we transform X with another distribution W, which is not the true distribution of the population, then $F_W(X)$ is not uniformly distributed. This observation provides a visual tool to compare different fits: indeed, suppose that sample X might have been sampled from a population of distribution Y or W. We may compute the two transformed samples $Z_1 = F_Y(X)$ and $Z_2 = F_w(X)$: if X was drawn from distribution Y, then the empirical distribution of Z_1 is likely to be close to the standard uniform, while if the true distribution was W, then we expect that Z_2, instead of Z_1, would be close to the standard uniform. A synthetic example of PIT is presented in Fig. 7. A random sample X of 10000 (non-censored) measures have been extracted from a standard normal distribution Y (Fig. 7A). The exponential Y_E models (hypotheses) has been fit with MLE to the random sample. Figures 7B and 7C represent the empirical frequency histogram and the empirical CDF (ECDF) of Z for both models ($Z_n = F_n(X)$ and $Z_e = F_e(X)$). Both figures show that the normal estimated model follows a standard uniform while the exponential model does not. The ECDF visualization of Fig. 7C is preferred over the simple frequency binning of Fig. 7B since multiple models can be tidily represented with the reference uniform distribution appearing as the diagonal of the plot. The closer the ECDF of Z is to the diagonal of the plot, the closer the fit is to the true underlying model. Values in the PIT plot that are below the diagonal indicate an overestimation of the model's CDF in respect to the empirical cumulative. Conversely values that are above the diagonal will result in an underestimation of the model's CDF. Hence the PIT provides a simple visual, yet powerful method to estimate which parametric model better fits the empirical data.

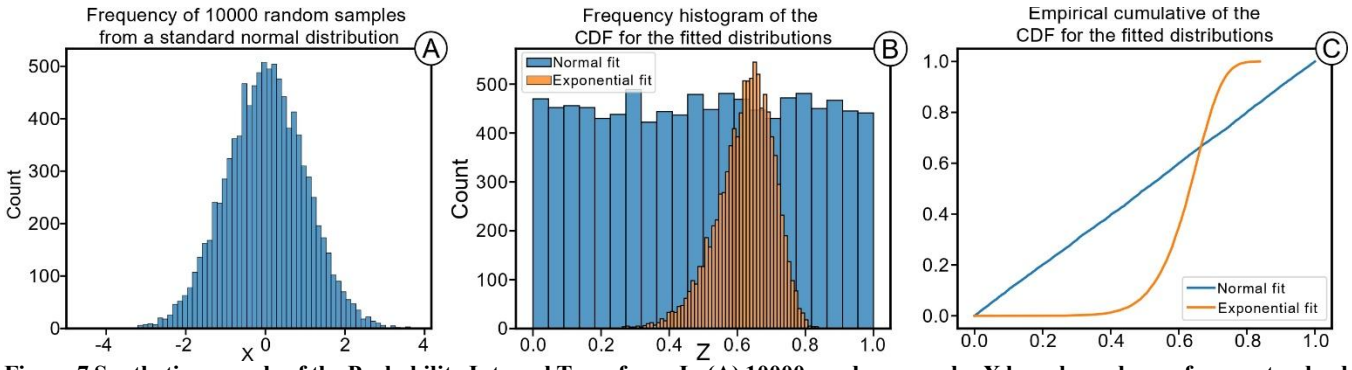


Figure 7 Synthetic example of the Probability Integral Transform. In (A) 10000 random samples X have been drawn from a standard normal distribution Y. In (B) the frequency distribution of Z from two different models (hypotheses): normal (Y_n) and exponential (Y_E). In (C) the empirical cumulative of Z for both models. From figure (B) and (C) it is possible to observe the effect of the Probability Integral Transform. In B the empirical frequency histogram of the normal model is remarkably close to a uniform distribution while the exponential model is not. This is visualized much more clearly in C where the normal model is the closest to the diagonal line y = x (i.e. the standard uniform).

3.5.2 A quantitative approach using distances

While the PIT visual approach is very intuitive, we propose to also use a quantitative approach, calculating four different statistics. The first is the Akaike Information Criterion (AIC, Akaike, 1974) that, by using the result of the MLE, quantifies the distance between the true natural phenomena and the estimated model as:

$$AIC = 2k - 2\ln\mathcal{L}(\hat{\theta}) \quad [6]$$

where k is the number of parameters used by the model (i.e. the dimension of the parameter vector) and $\hat{\theta}$ is the k-dimensional parameter vector that maximizes the likelihood of the model ($\mathcal{L}$). This formulation outputs a real number (either positive or negative) which should be small as the distance between the true population and the model decreases. If multiple models are tested, then it is possible and advised to calculate the Δ_i parameter i.e. the distances between the different models to the best scoring one (AIC_{min}):

$$\Delta_i = AIC_i - AIC_{min} \quad [7]$$

where i is the index of the proposed model.

From Δ_i it is possible to obtain the weight of evidence (w_i) of the given model with the following formula:

$$w_i = \frac{e^{-\Delta i/2}}{\sum_{r=1}^{R} e^{-\Delta r/2}} \quad [8]$$

where R is the total number of tested models (i.e. models deemed as reasonable by the researcher).

The weight of evidence outputs a value between [0; 1] that represents, in a set of proposed hypotheses, how likely it is that the model comes close to the true underlying process. The closer to 1 the more likely it is that, in the pool of candidates, the model represents the true underlying process (Burnham and Anderson, 2004). Since AIC and the derived formulas are directly based on the MLE, they are affected by the same limitations discussed in chapter 3.4 (Akaike, 1974). We thus propose to calculate three different distances between the model and the data (Kim, 2019). Usually, distances are calculated by comparing the empirical cumulative frequency with the cumulative frequency of the model; however, the data are censored and thus we use the empirical cumulative frequency estimated from Kaplan-Meier. Moreover, Kim, (2019) proposes to calculate the distances using the data transformed with PIT (i.e. $Z = F_y(X)$). We would like to highlight that purely under the point of view of the calculation, the distances are the same with or without using the transform, however with PIT the data are "normalized" and thus the different models are compared over the same scale (0, 1). Formula 3 then can be rewritten as:

$$\hat{G}_n = \begin{cases} 0 & , z < Z_1 \\ 1 - \prod_{Z_r \leq z} \left(\frac{N-r}{N-r+1}\right)^{\delta_r} & , z \leq Z_N \\ 1 & , z > Z_1 \end{cases} \quad [9]$$

where r are the indexes of each data point and δ_r is the same on-off switch of Eq. (5).

Kim (2019) then proposes to calculate:

- The Kolmogorov-Smirnov statistic (DC_n), representing the maximum distance between the two cumulative curves (Smirnov, 1939).
- The Koziol and Green statistic (Ψ^2), representing the sum of squared distances between the two cumulative curves (Koziol and Green, 1976).
- The Anderson-Darling statistic (AC_n^2), representing the weighted sum of squared distances between the two cumulative curves, imparting more weight than the Kolmogorov-Smirnov on tail observation (the closer to 0 or 1 the higher the weight; Anderson and Darling, (1954)).

The three statistics can be modified to accommodate the presence of censored data by using the Kaplan-Meier estimator.

The Kolmogorov-Smirnoff statistics is generally calculated as:

$$DC_n = max(DC_n^-, DC_n^+) \quad [10]$$

where with censored data:

$$DC_n^+ = max\left(\hat{G}_n(Z_r) - Z_r\right) \quad [10a]$$

$$DC_n^- = max\left(Z_{r+1} - \hat{G}_n(Z_r)\right) \quad [10b]$$

The Koziol-Green statistics is a generalization of the Cramer-von Mises statistics (Koziol and Green, 1976):

$$\Psi^2 = N \int_0^1 \left(\hat{G}_n(z) - z\right)^2 dz \quad [11]$$

Which considering censored data can be written as (Appendix 2 in Koziol and Green (1976)):

$$\Psi^2 = \frac{1}{3}N + N \sum_{r=1}^{N} \hat{G}_n(Z_n) \times (Z_{r+1}) - Z_r \times \left[\hat{G}_n(Z_r) - (Z_{r+1} - Z_r)\right] \quad [11a]$$

Finally, the Anderson-Darling distance (Anderson and Darling, 1954), generally defined as:

$$AC_n^2 = N \int_0^1 \frac{(\hat{G}_n(z) - z)^2}{z(1-z)} dz \quad [12]$$

where $^{1}/_{z(1-z)}$ is the weight, can be calculated accounting for censoring as a finite sum (Kim, 2019):

$$AC_n^2 = N \sum_{r=1}^{N-1} \left(\hat{G}_n(Z_r)\right)^2 \times (-\ln(1-Z_{r+1}) + \ln(Z_{r+1}) + N \times (1-Z_r) - \ln(Z_r)) -$$
$$2N \sum_{r=1}^{N-1} \left(\hat{G}_n(Z_r)\right) \times (-\ln(1-Z_{r+1}) + \ln(1-Z_r)) - N \times \ln(1-Z_N) + N \times \ln(Z_N) + N \quad [12a]$$

Once these distances are calculated for each model, we propose to rank the models independently for each distance (from minimum to maximum). If all distances converge (i.e. for the same model the same rank is assigned) then we consider this as sufficient proof for the overall ranking of the model in the list. By comparing multiple rankings, even if the calculated distances are ranked in differently, it is still possible to make a sensible guided choice, for example by using the PIT representation together with the mean ranking position or following a specific type of distance.

4 Case studies

We demonstrate the applicability of the discussed theory by using the FracAbility package on three different case studies. The first case study focuses on the characterization of a very regular and densely fractured fracture network with a simple boundary geometry. The second focuses on analyzing a typical average-quality outcrop, with multiple fracture sets and complex boundary geometry. Finally, the third case study presents an application with spacing analysis to demonstrate that any type of length can (and must) be corrected from right-censoring bias. We propose to test six statistical models of length, frequently adopted in literature (Bonnet et al., 2001) or used in stochastic DFNs:

- Lognormal
- Truncated power law (power law in the following)
- Normal
- Weibull
- Exponential
- 2-parameters Gamma (gamma in the following)

4.1 First case study: Cava Pontrelli, Puglia (IT)

This first case study focusses on the analysis of a single NW-SE striking fracture set present in an abandoned quarry near the town of Altamura in Puglia (Italy). The study area is located in the Apulian Platform (Panza et al., 2019),

representing the forebulge of the southern part of the Apennines fold and thrust belt. The outcrop is characterized by an extensive horizontal pavement of about 18.000 m^2, showing densely fractured Cretaceous platform limestone of the Altamura Limestone Formation (Panza et al., 2016). The continuous maintenance that followed the discovery of thousands of dinosaur footprints on the quarry pavement (Nicosia et al., 1999), made it possible to obtain an exceptionally clean outcrop surface. This in turn resulted in the definition of a simple boundary geometry (with a couple of interpretational holes) and the digitalization of a very regular and dense fracture network (1941 fractures). The combination of these factors led to a very low percentage of censoring with only 8.9% of the total fractures being censored (Fig. 8).

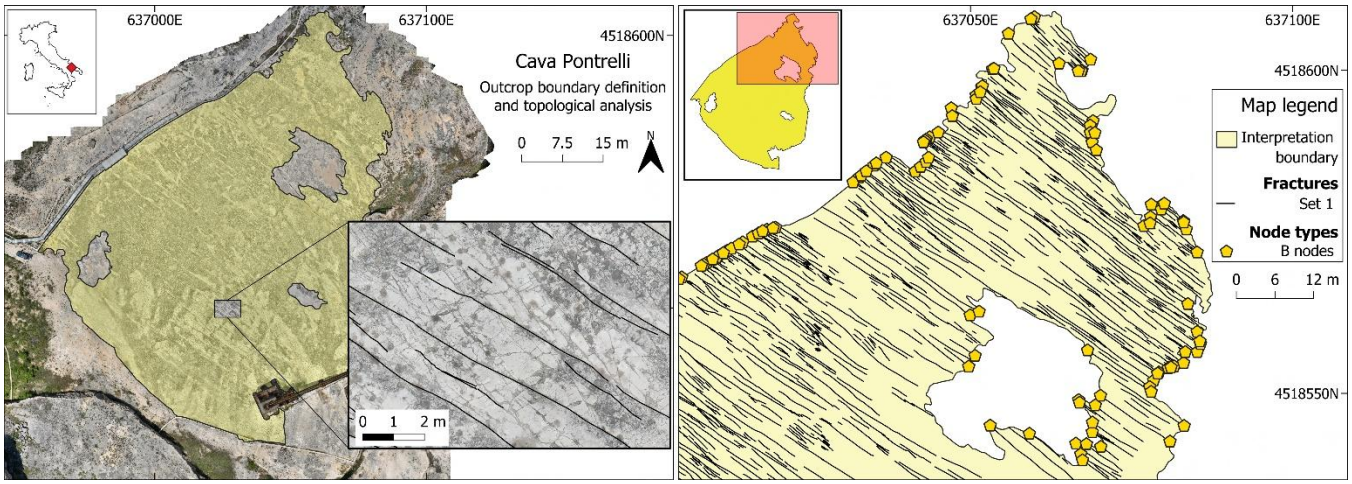


Figure 8 Overview maps of the first case study area. Pictured on the left, the general overview of the area with the boundary geometry overlayed on the orthophoto and a small sample of the digitized facture set. Pictured on the right, a subarea of the quarry with the boundary, fracture traces and boundary nodes (yellow pentagons).

We applied survival analysis as implemented in FracAbility, and in Fig. 9 the resulting PIT plot is shown. From this, the most representative model appears to be the lognormal, confirmed by all the distances calculated in Table 1 (ordered by ascending Akaike scores) (see Fig. 10 for a summary of the fit). Figure 9 shows that the PIT of the lognormal model is quite linear with gentle undulations indicating a slight underestimation for lengths between 1m and 2m (accounting for 20% of the measures) and slight overestimation for 3.5m and 6.5m (accounting for another 20% of the measures). For the calculated distances scores, the lognormal is followed by the gamma and Weibull models, however the rank scores of these last two models are not uniform. While the Koziol and Green distance favors the gamma model, the Anderson-Darling distance favors the Weibull model. Finally, the power law and normal models rank lowest in comparison to the other models, indicating that they are less able to represent the dataset.

Model	AIC	Δ_i	w_i	DC_n	Ψ^2	AC^2	AIC rank	DC_n rank	Ψ^2 rank	AC^2 rank	Mean rank
Lognormal	8522.15	0	1	0.02	0.07	0.56	1	1	1	1	1
Gamma	8742.2	220.05	0	0.07	2.97	19.53	2	4	4	3	3
Weibull	8770.37	248.22	0	0.06	2.53	19.18	3	2	3	2	3
Exponential	8774.85	252.7	0	0.06	2.33	19.53	4	3	2	4	3
Power law	10639.54	2117.4	0	0.32	71.71	354.2	5	6	6	6	6
Normal	10682.11	2159.97	0	0.18	24.54	148.01	6	5	5	5	5

Table 1 Models' distance tables and ranking scores tables for the Pontrelli quarry. The closer to 0 the better. For this dataset the lognormal is the most representative of the data in all the different distances while the power law and normal models are the least representative. The positions of the other models in between are less certain (especially the gamma and exponential).

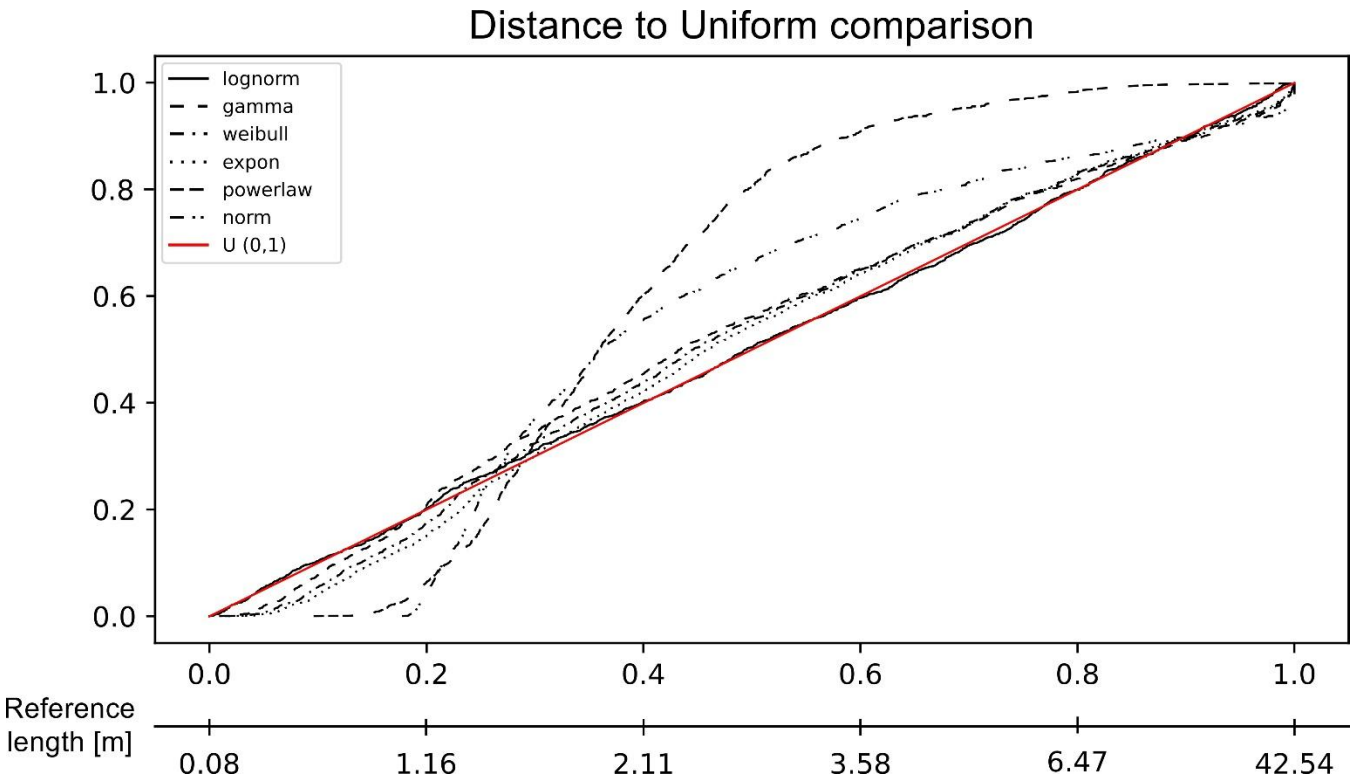
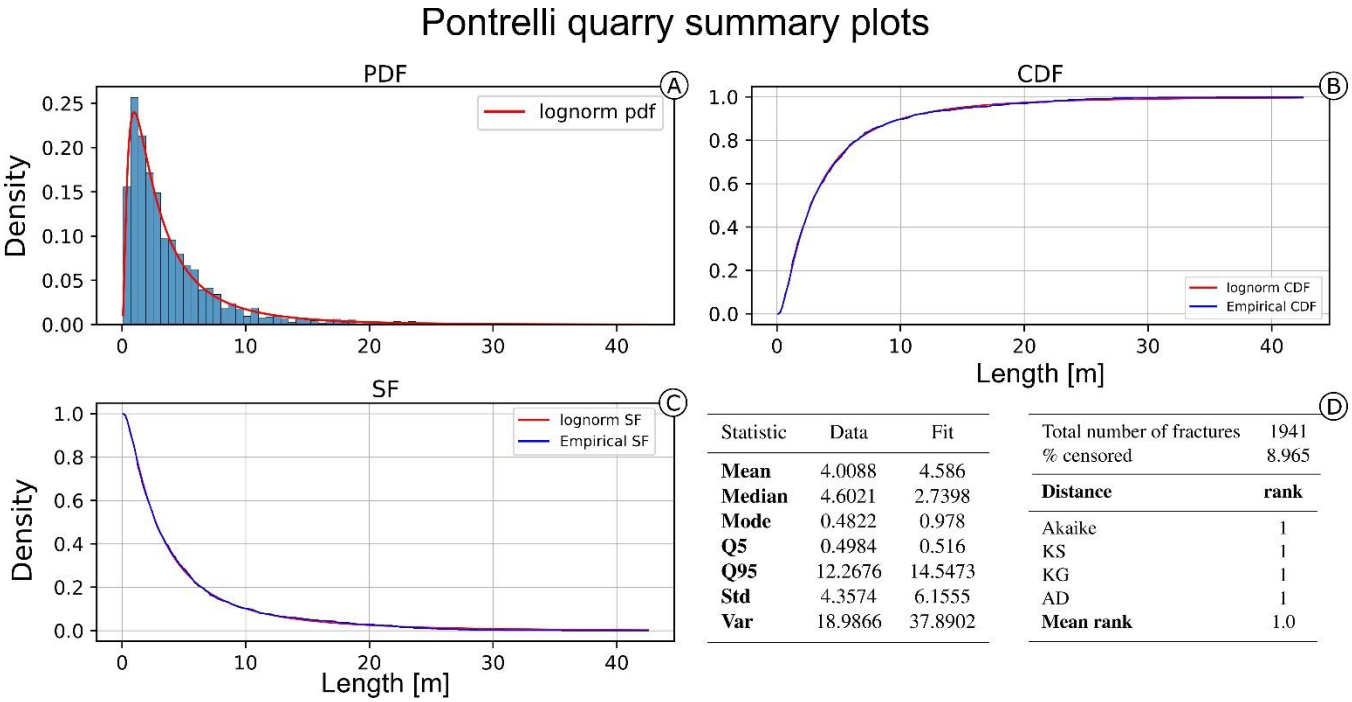


Figure 9 PIT visualization for the proposed length models in the Pontrelli quarry. In red represented the reference U(0,1), the closer the model line to this reference line the more representative the model. For this dataset, the lognormal model is the most representative following almost perfectly the reference line with some minor underestimation between 1.16m and 2.11m and overestimation between 3.58m and 6.47m.



Statistic	Data	Fit
Mean	4.0088	4.586
Median	4.6021	2.7398
Mode	0.4822	0.978
Q5	0.4984	0.516
Q95	12.2676	14.5473
Std	4.3574	6.1555
Var	18.9866	37.8902

Total number of fractures	1941
% censored	8.965
Distance	**rank**
Akaike	1
KS	1
KG	1
AD	1
Mean rank	1.0

Figure 10 Summary of the best fitting model (lognormal) for the Pontrelli quarry dataset. (A) The Probability Density Function against the histogram of the dataset. (B) The Cumulative Density Function and the (C) Survival Function against the empirical counterparts calculated with the Kaplan-Meier estimator. (D) The summary table of the main statistics for the data (e.g. sample mean, sample variance etc.) compared with the estimated model.

4.2 Second case study: Colle Salza, Valle d'Aosta (IT)

The second case study focuses on the analysis of a less ideal, albeit more realistic outcrop. The study area is located in the basement of the Western Alps, on paragneiss of the Monte Rosa Nappe (Piaz et al., 2003). The outcrop is cross-cut by several brittle fractures, Tertiary in age (Bistacchi and Massironi, 2000), and is characterized by a main central area of 1234 m^2 and two secondary smaller satellite areas of 9 m^2 and 3 m^2. The boundary geometry is highly concave thus leading to a high censoring fraction of fracture traces. Moreover, the outcrop is exposed on top of a small topographical height with a slightly ellipsoidal shape due to glacial erosion (i.e. a roche moutonnee) with the main axis directed NW-SE. This leads to an inevitable deformation of the orthophoto (and thus of the measured lengths) along the extremities of the analyzed area. The analyzed fractures are subdivided in two main sets, striking NE-SW (Set 1) and NW-SE (Set 2), conforming to the general trend of the area for brittle deformation (Bistacchi et al., 2000; Bistacchi and Massironi, 2000; Bistacchi et al., 2001).

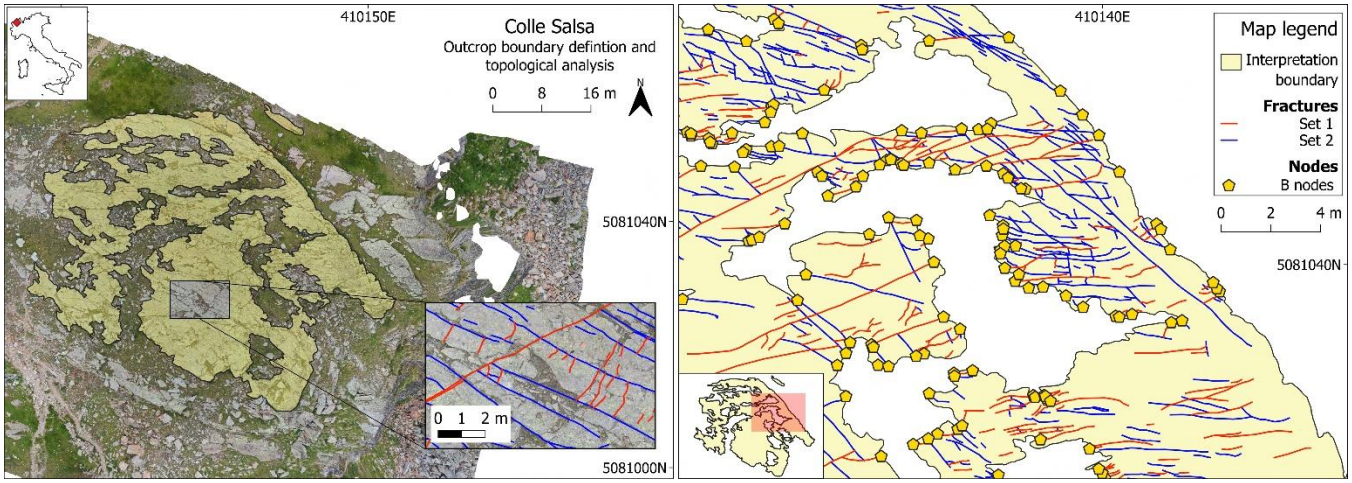


Figure 11 Overview maps of the second case study area. Pictured on the left, the general overview of the area with the boundary geometry overlayed on the orthophoto and a small sample of the digitized facture sets. On the right it is represented a subarea of the quarry with the boundary, fracture traces and intersection nodes between the two (yellow pentagons).

The total number of fractures is 1718 (24.7% censored) while set-wise the number of fractures is 692 (24.3% censored) for Set 1 and 1026 (24.9% censored) for Set 2 (Fig. 11). The different estimated models are represented for both sets in Fig. 12. In both cases the lognormal model, overall, better fits the data. Nonetheless, the fit of estimated models is visibly worse than in the first case study, particularly for Set 1 (Fig. 12A). The lognormal fit of Set 1 underestimates lengths between 0.34 m and 1.5 m (about 50% of the measures) while lengths above this value are generally overestimated. For the lognormal fit of Set 2 (Fig. 12B), there is a less relevant underestimation of the length values between 0.44 m and 2.57 m (about 60% of the measures) and a slight overestimation for longer length values. The distances values in Table 2 (ordered by ascending values of Akaike) confirm the lognormal as the most representative for both sets (see Fig. 13 for the summary plots of the best fit). For the other models, the mean rank value can be used to assign a final ranking. The mean ranks show that the gamma distribution is less representative than both the exponential and the Weibull (respectively the second and third most representative models).

Set 1											
Model	**AIC**	$\boldsymbol{\Delta_i}$	$\boldsymbol{w_i}$	$\boldsymbol{DC_n}$	$\boldsymbol{\Psi^2}$	$\boldsymbol{AC^2}$	**AIC rank**	$\boldsymbol{DC_n}$ **rank**	$\boldsymbol{\Psi^2}$ **rank**	$\boldsymbol{AC^2}$ **rank**	**Mean rank**
Lognormal	1118.98	0.00	1.00	0.06	0.55	5.05	1	1	1	1	1
Gamma	1264.59	145.61	0.00	0.12	2.67	23.59	2	4	4	4	3.5
Exponential	1280.38	161.40	0.00	0.10	2.32	20.82	3	2	2	2	2.25

Weibull	1281.01	162.03	0.00	0.11	2.40	21.63	4	3	3	3	3.25
Power law	1935.96	816.98	0.00	0.36	32.68	160.01	5	6	6	6	5.75
Normal	2068.24	949.26	0.00	0.22	11.84	79.92	6	5	5	5	5.25
Set 2											
Model	**AIC**	$\boldsymbol{\Delta_i}$	$\boldsymbol{w_i}$	$\boldsymbol{DC_n}$	$\boldsymbol{\Psi^2}$	$\boldsymbol{AC^2}$	**AIC rank**	$\boldsymbol{DC_n}$ **rank**	$\boldsymbol{\Psi^2}$ **rank**	$\boldsymbol{AC^2}$ **rank**	**Mean rank**
Lognormal	2204.06	0.00	1.00	0.04	0.39	4.53	1	1	1	1	1
Gamma	2379.13	175.07	0.00	0.09	2.97	29.94	2	4	4	4	3.5
Exponential	2390.81	186.75	0.00	0.09	2.54	25.31	3	2	2	2	2.25
Weibull	2392.34	188.28	0.00	0.09	2.62	26.24	4	3	3	3	3.25
Power law	3109.95	905.89	0.00	0.31	35.29	167.56	5	6	6	6	5.75
Normal	3421.72	1217.66	0.00	0.21	15.24	137.05	6	5	5	5	5.25

Table 2 Models' distance and rank tables for both Colle Salza sets. The closer to 0 the better. For this length dataset, the lognormal is the most representative of the data for all the different distances. Using the mean rank, the lognormal is followed by the exponential, Weibull and gamma distributions. As with the first case study the Power law and normal distribution are the least representative.

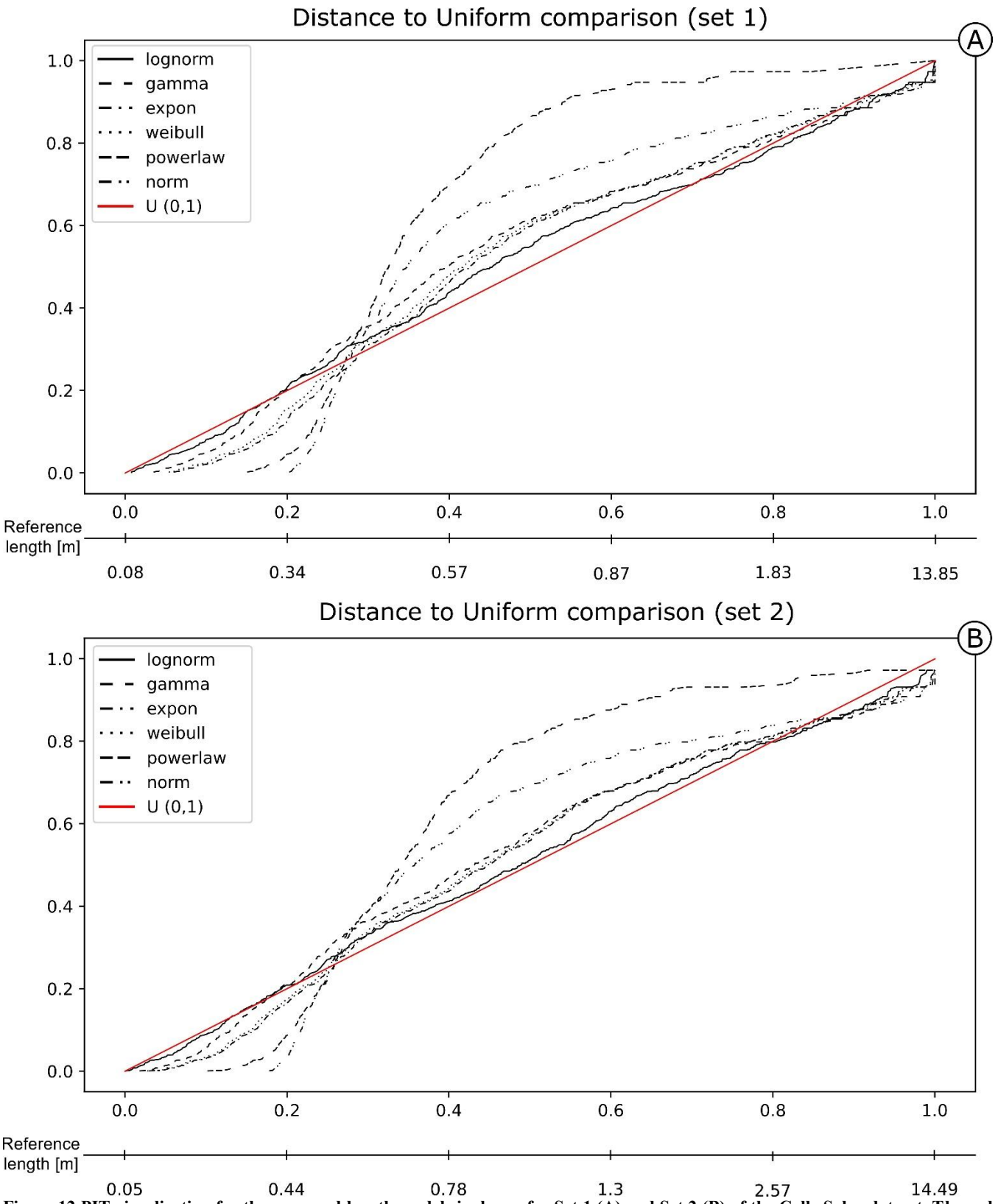
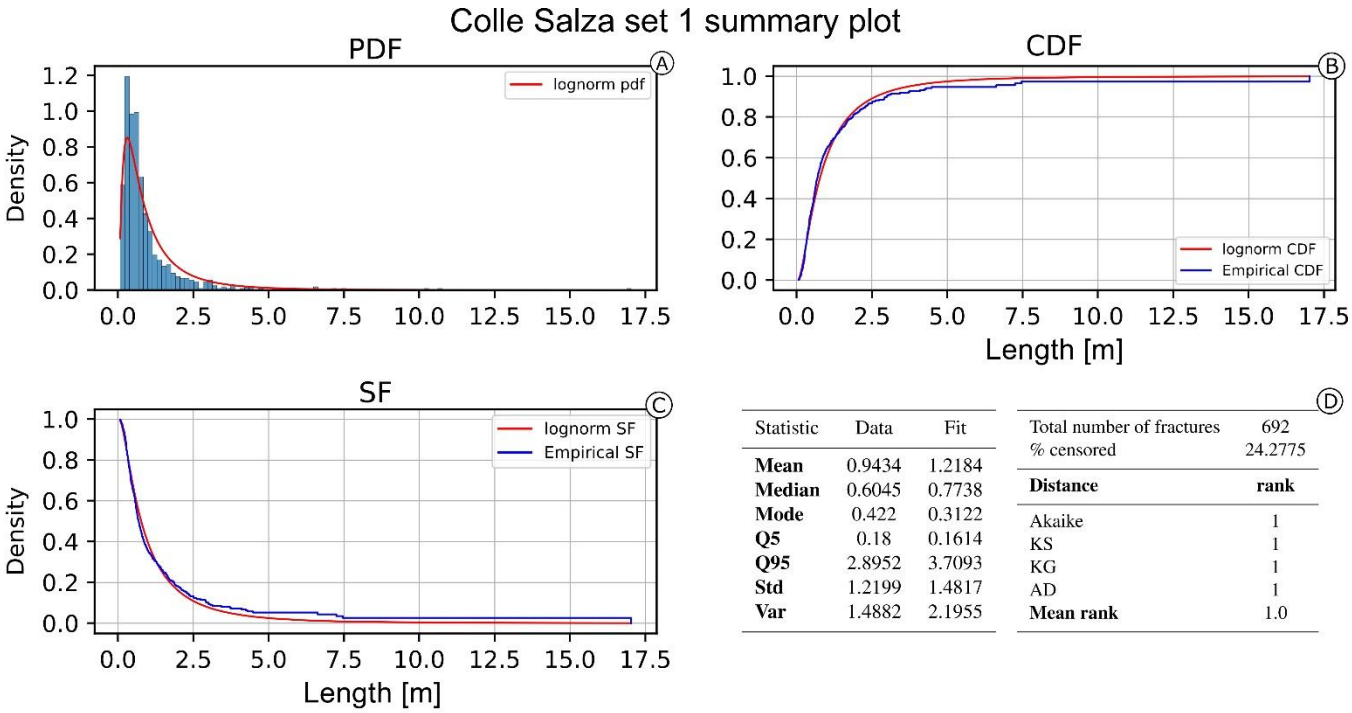


Figure 12 PIT visualization for the proposed length models is shown for Set 1 (A) and Set 2 (B) of the Colle Salza dataset. The red line represents the reference U(0,1); the closer a model's line is to this reference, the more representative the model. Among the models, the lognormal distribution demonstrates the closest fit to the reference line in both sets, although showing a worst fit to that

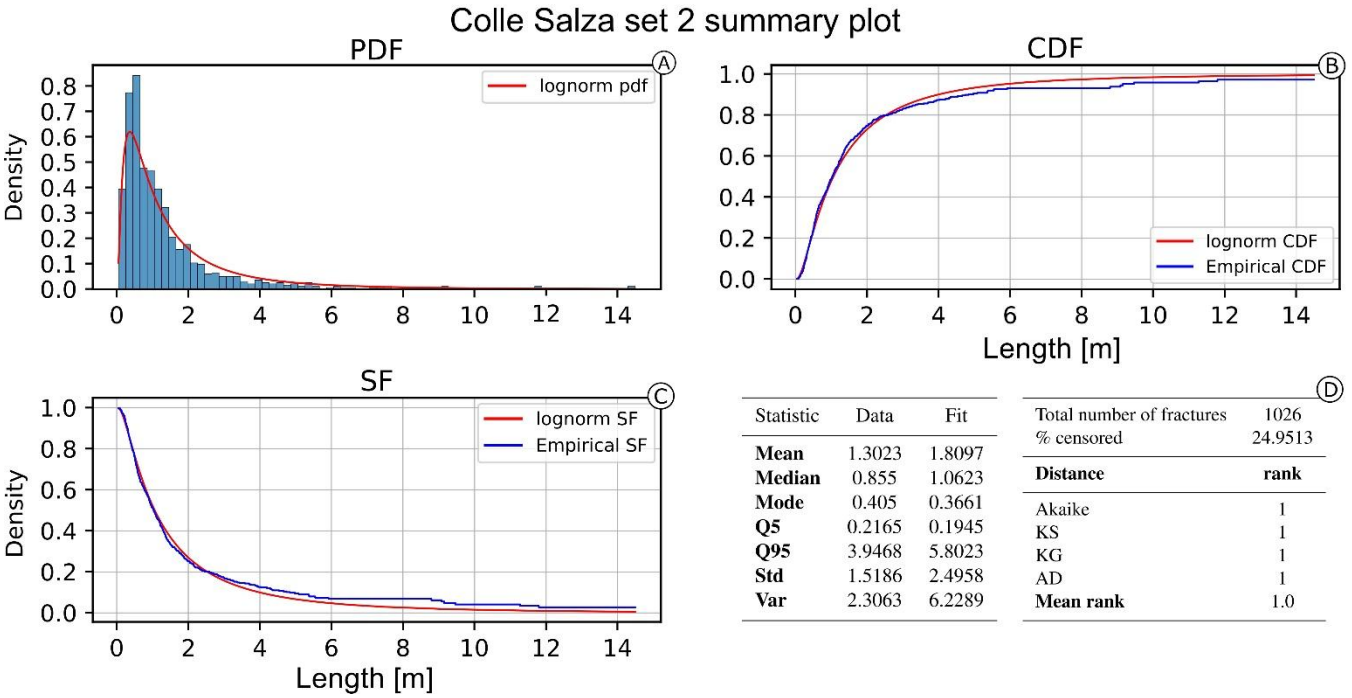
observed in the first case study. Across both sets, all estimated models exhibit less linearity, with notable underestimation between 0.34 m and 1.5 m in Set 1 (A) and between 0.44 m and 2.57 m in Set 2 (B).



Statistic	Data	Fit
Mean	0.9434	1.2184
Median	0.6045	0.7738
Mode	0.422	0.3122
Q5	0.18	0.1614
Q95	2.8952	3.7093
Std	1.2199	1.4817
Var	1.4882	2.1955

D

Total number of fractures	692
% censored	24.2775
Distance	**rank**
Akaike	1
KS	1
KG	1
AD	1
Mean rank	1.0



Statistic	Data	Fit
Mean	1.3023	1.8097
Median	0.855	1.0623
Mode	0.405	0.3661
Q5	0.2165	0.1945
Q95	3.9468	5.8023
Std	1.5186	2.4958
Var	2.3063	6.2289

D

Total number of fractures	1026
% censored	24.9513
Distance	**rank**
Akaike	1
KS	1
KG	1
AD	1
Mean rank	1.0

Figure 13 Summary of the best fitting model for the Colle Salza dataset (Set 1 and 2). (A) The probability density function against the histogram of the dataset. (B) The Cumulative Density Function and the (C) Survival Function against the empirical counterparts calculated with the Kaplan-Meier estimator. (D) The summary table of the main statistics for the data (e.g. sample mean, sample variance etc.) and estimated model.

4.3 Third case study: Spacing analysis

Survival analysis can be used also to analyze the spacing length distribution for each fracture set. Thus, the same workflow was applied to spacing measurements of both Set 1 and Set 2 of the Colle Salza dataset. It is worth noting that the censoring analysis is a secondary part in the analysis for spacing. Analyzing the spatial arrangement of the fractures in the network (such as Marrett et al. (2018) and Bistacchi et al. (2020)) is of fundamental importance. However, we decided not to include this analysis and focus mainly on censoring to avoid increasing the length of an already dense text. Spacing is defined as the distance between two fractures of the same set, measured perpendicular to the average fracture plane attitude (Bistacchi et al. (2020) and refs. therein). Traditionally, this type of statistic is obtained in the field with scanline surveys, with properly oriented scanlines or by applying the Terzaghi correction (Terzaghi, 1965). With DOMs, this procedure can be easily sped up with custom scripts, programmatically tracing a large number of parallel scanlines and thus increasing the number of spacing measurements (De Toffoli et al., 2021; Storti et al., 2022). However, this leads to a higher number of censored spacing measurements occurring at the outcrop boundary and where "holes" are present. As with standard lengths, the introduction of a boundary affects the actual estimation of the spacing by a degree depending on the outcrop's convexity (Fig. 14). Perfectly convex boundaries (Fig. 14B) will lead to minor censoring, affecting only the ends of the scanlines, while more realistic non-convex boundaries will increase the censoring percentage because the boundary will cross in multiple points the same scanline (Fig 14C).

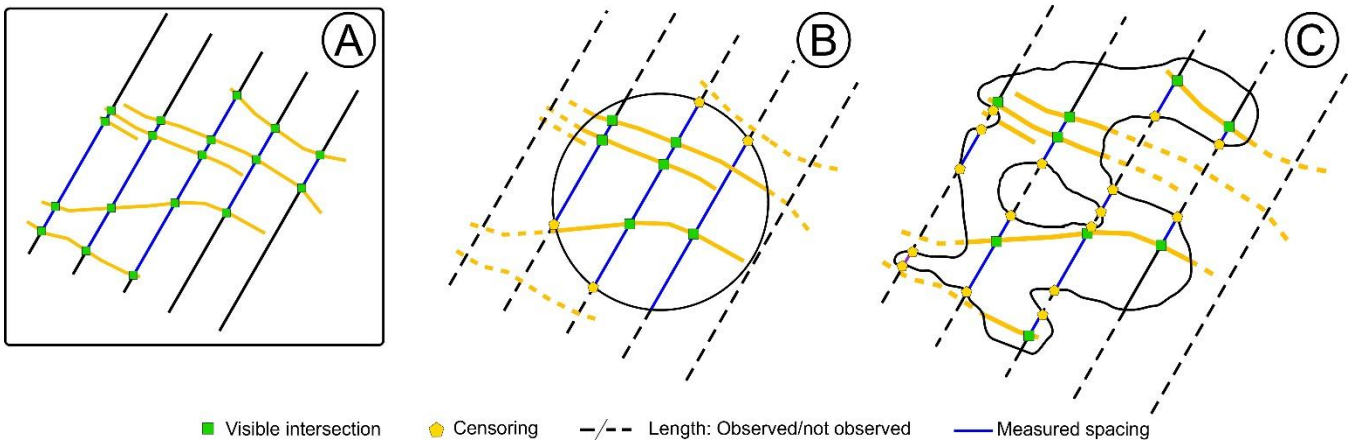


Figure 14 The effect of boundary geometry on censoring for spacing distribution modelling. (A) The ideal situation where no censoring occurs. Colored in green are the visible intersections between the scanline and the fracture while in blue the measured spacing segment. (B) The effect of a perfectly convex boundary (i.e. a circle) on the estimation of censoring. For the "visible" scanlines only the final segments are censored. (C) The effect of a complex boundary geometry. Here the number of censoring measurements increases drastically and that censoring it is no longer limited to the ends of the scanlines.

For the Colle Salza dataset, 1282 and 1119 spacing measurements have been obtained for fracture sets 1 and 2 respectively (with 52.5% and 36.5% censoring). Figure 15 shows the distances of the different models for the obtained spacing values. In this case, for Set 1, an extreme underestimation of the lognormal model is present for values longer than 4 m (Fig 15A), while for the other proposed models, particularly the Weibull, the distances at the same values are smaller. This is visible also in the distances tables (Table 3). Although the Akaike distance for the lognormal model is the smallest, the difference from the other models (Δ_i) is not as big as the other examples. Thus, the weight of evidence (w_i) also is not completely in favor of the lognormal model. The poor fit of the lognormal model is also confirmed by the other distances that rank the lognormal lower than the Weibull and Gamma models (ranking first and second respectively). All these factors point to the fact that in this case the Weibull is the most representative model (see Fig. 16 for a summary representation of the best fit). Conversely, the lognormal model is clearly the most representative for Set 2 (Fig 15B). The PIT plot showing a quite linear behavior and all the calculated distances are ranking first.

Set 1											
Model	**AIC**	$\boldsymbol{\Delta_i}$	$\boldsymbol{w_i}$	$\boldsymbol{DC_n}$	$\boldsymbol{\Psi^2}$	$\boldsymbol{AC^2}$	**AIC rank**	$\boldsymbol{DC_n}$ **rank**	$\boldsymbol{\Psi^2}$ **rank**	$\boldsymbol{AC^2}$ **rank**	**Mean rank**
Lognormal	2508.66	0.00	0.55	0.11	1.25	13.32	1	4	3	4	3
Weibull	2509.59	0.93	0.35	0.05	0.75	6.17	2	1	1	1	1.25
Gamma	2512.84	4.17	0.07	0.06	1.03	8.37	3	2	2	2	2.25
Exponential	2514.14	5.48	0.04	0.07	1.65	13.27	4	3	4	3	3.5
Power law	2682.15	173.49	0.00	0.16	12.65	116.42	5	6	6	5	5.5
Normal	3247.15	738.49	0.00	0.14	10.45	169.38	6	5	5	6	5.5
Set 2											
Model	**AIC**	$\boldsymbol{\Delta_i}$	$\boldsymbol{w_i}$	$\boldsymbol{DC_n}$	$\boldsymbol{\Psi^2}$	$\boldsymbol{AC^2}$	**AIC rank**	$\boldsymbol{DC_n}$ **rank**	$\boldsymbol{\Psi^2}$ **rank**	$\boldsymbol{AC^2}$ **rank**	**Mean rank**
Lognormal	1652.30	0.00	1.00	0.03	0.15	1.47	1	1	1	1	1
Gamma	1709.71	57.41	0.00	0.07	1.40	23.53	2	4	4	4	3.5
Exponential	1712.78	60.48	0.00	0.06	1.18	19.66	3	2	2	2	2.25
Weibull	1714.62	62.31	0.00	0.06	1.22	20.53	4	3	3	3	3.25
Power law	2340.23	687.93	0.00	0.31	38.10	180.73	5	6	6	6	5.75

Normal	2603.07	950.77	0.00	0.17	11.56	136.26	6	5	5	5	5.25

Table 3 Models' distance and rank tables for the spacing calculated on both Colle Salza sets. The closer to 0 the better. For Set 1 the low performance of the lognormal in the estimation of longer values (Fig. 15A) is visible in all distances. Even if the Akaike distance calculated for the lognormal is the smallest, the difference from the other calculated values is not enough to completely justify the selection of this model. This is also shown in the other distances where the lognormal clearly underperforms. On the other hand, for Set 2 the lognormal model is the most representative for the spacing values for all the different distances.

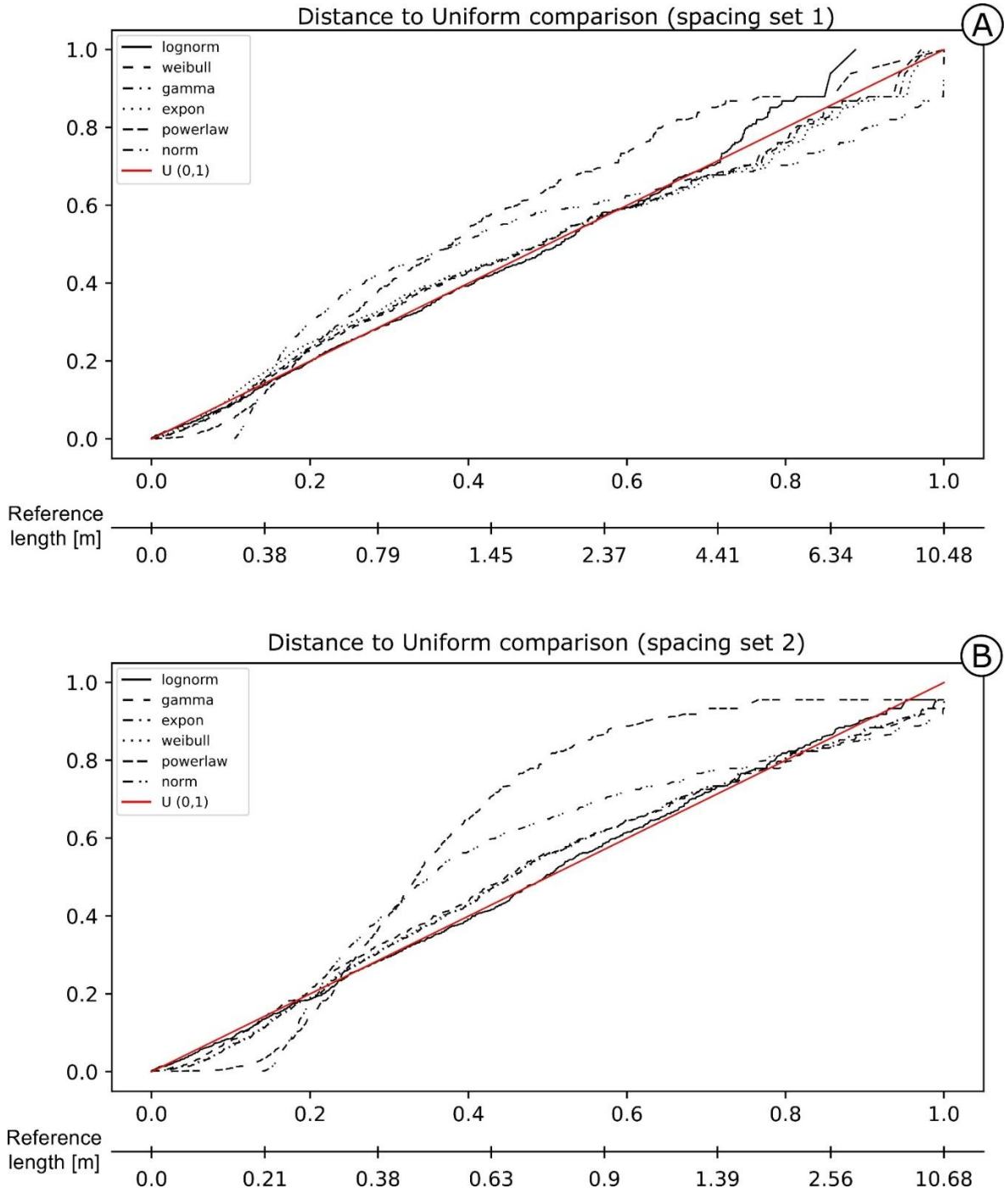


Figure 15. PIT visualization for the proposed length models is shown for Set 1 (A) and Set 2 (B) of the Colle Salza dataset. The red line represents the reference U(0,1); the closer a model's line is to this reference, the more representative the model. In contrast with

the other case studies, the lognormal distribution demonstrates a marked underestimation for values longer than 4.4m for the spacing of Set 1 (A). In this case the closest fit is the Weibull. Conversely for the spacing of Set 2 (B), the lognormal model is again the model closest to the reference line, performing similarly to the first case study.

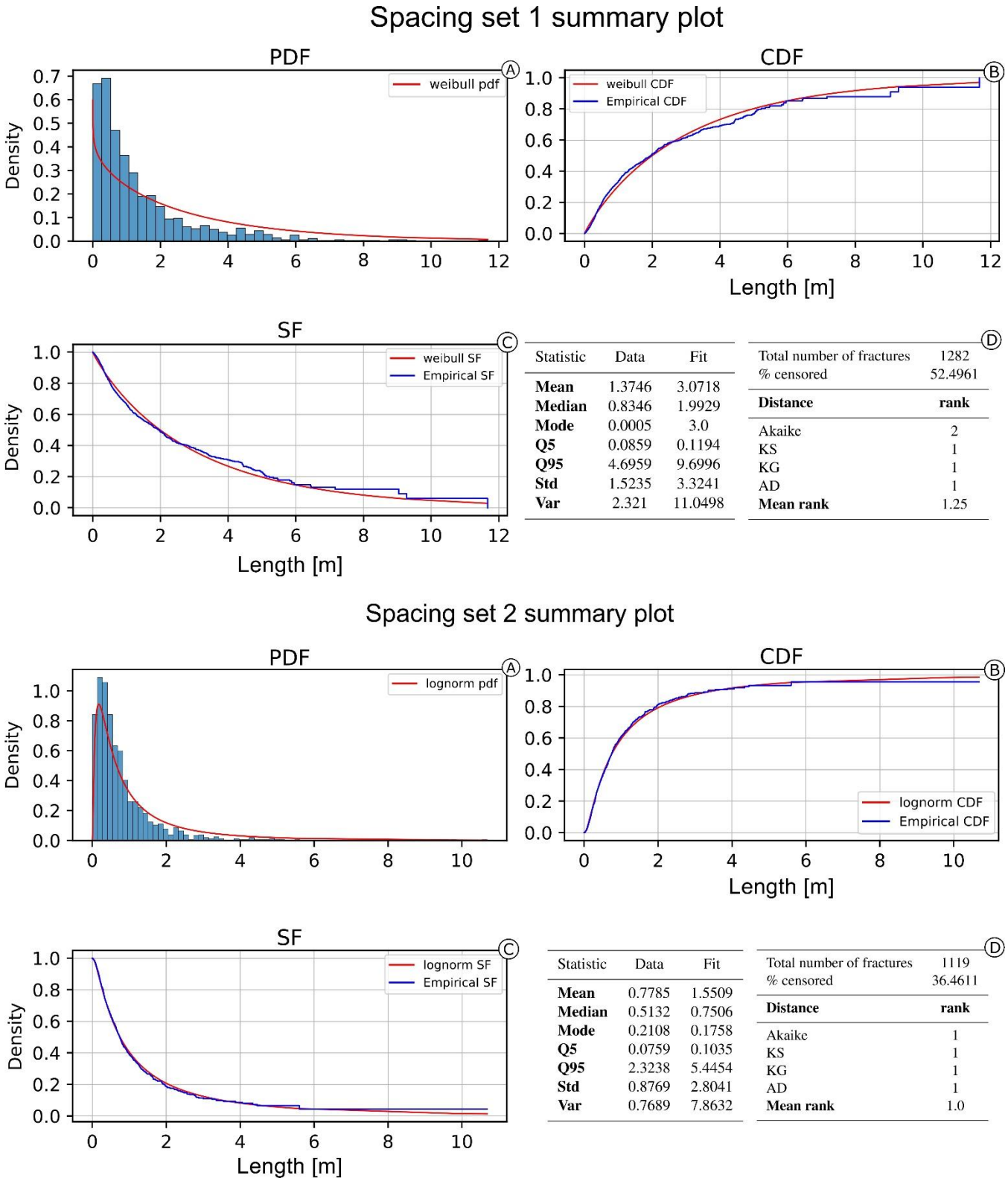


Figure 16 Summary of the best fitting model for the spacing dataset (Set 1 and 2). (A) The probability density function against the

histogram of the dataset. (B) The Cumulative Density Function and the (C) Survival Function against the empirical counterparts calculated with the Kaplan-Meier estimator. (D) A summary table of the main statistics for the data (e.g. sample mean, sample variance etc.) and estimated model. The spacing of Set 1 is the only case where the data is not lognormally distributed and the frequency histogram reflects such conclusion.

5 Discussion

With this work we delineate a robust statistical framework to quantitively analyze and statistically model fracture trace length and spacing. The statistical distributions are estimated using the entire length of fracture traces, following the original works of Baecher and Lanney (1978) and Baecher (1980), and not the length of branches as defined in Sanderson and Nixon (2015). Branches offer a useful topological abstraction of the network (making it possible to classify node intersections), however they do not carry a real geological or physical meaning, and thus a distribution obtained by fitting branch-length will have a different meaning compared to one obtained from complete lengths. Furthermore, length can sometimes be affected by a subjective bias leading to the uncertainty of which structure should be measured. For example, with segmented or en echelon fractures do we measure the single segments or consider their cumulative sum? By approaching this problem with branches, a clear objective answer cannot be reached while by considering a 2D lineament as a representation of the 3D geological structure, contextualized in the geological history of the outcrop, a more rational and informed decision can be made (i.e. Forstner and Laubach, 2022).

The first and crucial point of discussion is the objective of having a representative unbiased statistical model that is sufficiently likely to reproduce the observed length data. In particular, we want to stress that there is no real interest, nor practical reason, and no theoretical possibility to find the real underlying distribution, while there is a strong necessity to fit specific simplified parametric models, useful for stochastic modelling applications (such as DFNs). Due to this, non-parametrical methods such as those proposed by Mauldon et al. (2001) are unfit since they are not linked to any model. Furthermore, having an unbiased statistical model can be also extremely important for engineering applications. For example, Pahl (1981) state at the end of his work that in geomechanical application it could be useful to know the longest trace likely to occur in a group of traces. With a parametric model this can be easily estimated by checking the length values associated to a probability chosen depending on the safety margin that is needed for the use case. Modern alternatives that use a simple implementation of MLE such as FracPaQ (Healy et al., 2017) are a good step forward, however the censoring bias is still present and thus the results are biased, always underestimating length. The approach proposed here provides for the middle ground that was missing. We have shown that it is possible to use survival analysis to model censored length datasets and we have showcased a new approach in model selection via quantitative distances to rank and choose the most representative model from a pool of simple candidates that can be used in common modelling applications.

As stated, to use MLE for a given model we assumed that, in geological applications, the only censoring mechanism present is random censoring. Both left and interval censoring, by definition, cannot happen because the event of

interest coincides with the end of the fracture. Thus, for left censoring, fractures terminate outside the boundary and thus cannot be measured (this is analogous patients never entering the study because they are already dead). The same applies with interval censoring that could be erroneously identified when holes are present in the sampling area or when boundaries are concave. However, since the measured event is the end of the fracture, on the other side of the interruption no fracture should be present. In terms of classic survival analysis theory, this would compare to a patient's loss of follow-up in the study period, caused by voluntary or involuntary exit from the study group, thus classifying it again as right censoring (Leung et al., 1997).

To correctly classify censoring as random, we must assume independence between the censoring and length distribution. By "independence" it is intended that the mechanisms behind the generation of a fracture length distribution is different from the mechanisms that censors such lengths. The boundary, which represents censoring, is usually the product of secondary events that occur after fracture genesis (i.e. alteration, debris hiding part of the outcrop, vegetation, human activity, etc.). Thus, albeit it is often the case that such events are controlled by preexisting structures, the physical processes that caused censoring are not the same that generated the fracture set and thus the original length distribution. This leads to an important implicit caveat where the measured lengths must be related only to the mechanism that we are interested in modelling, for example lengths that are surely linked to tectonics and no other secondary events. Such discussion highlights that the assumption of independence is difficult to rigorously prove since the true distribution of the length of fractures is not known (we only observe a set of complete and censored data). In some applications (Eppes et al., 2024), this assumption may not hold, and a more in-depth study may be required to prove the independence hypothesis before proceeding. Nonetheless, we believe that it can be safely assumed in geological applications when the appropriate field work and *a posteriori* analysis are carried out.

As a second point, we would like to discuss is the relationship between the fraction of censored data and the uncertainty in the estimated statistical model. There is not much literature on this relationship probably because, since traditional survival analysis is applied in function of time, the time boundary can be expanded if necessary. For example, if a study shows a censoring rate that is deemed too high, the researcher could repeat the experiment and simply extend the study period to observe the event of interest. In our case, the spatial boundary usually cannot be expanded and thus knowing the effects of censoring on estimation is an important aspect. For example, it would be useful to know how censoring influences the estimations based on simplistic approaches described in section 3, or the censoring percentage value above which survival analysis must be used, or the value above which even survival analysis fails. Moreover, it can be also useful to estimate the precision of estimation with survival analysis, knowing the censoring percentage. The case studies discussed in this work show a censoring percentage ranging from 8.96% (first case) to 52.50% (third case), and effectively showcase the influence of censoring on the estimation. For each dataset it is possible to use the best fitting statistical model defined in the results and use PIT

plots to visualize the differences. The comparison with the circular scanline approaches was intentionally omitted because it cannot be compared in any way. Being a non-parametrical approach, cumulative frequencies cannot be compared and comparing mean values without a variance is not informative. Figure 17 shows that in all cases, the estimated model using survival analysis closely follows the reference uniform, while the other two approaches (i.e. ignoring censoring and removing censored data) always overestimate the cumulative frequency, diverging with a rate positively correlated with the censored fraction. This translates in an underestimation of a model's parameters proportional to the censoring percentage. This effect is due to the fact that measured lengths of censored fractures will always be shorter than their true lengths. Thus, by using the first simple approach, the dataset is essentially "polluted" by shorter fractures thus always decreasing the measured mean. The second simple method will also lead to an underestimation of the mean because of the size bias. However, this second method can be less impacted by censoring. For example, if a fracture population has a very small standard deviation (i.e. almost all fractures have the same length) and/or fractures are occurring in an outcrop that is much bigger than the characteristic fracture length, then removing censored values would not have a great impact on the estimation. But, even if small, the underestimation will always be present.

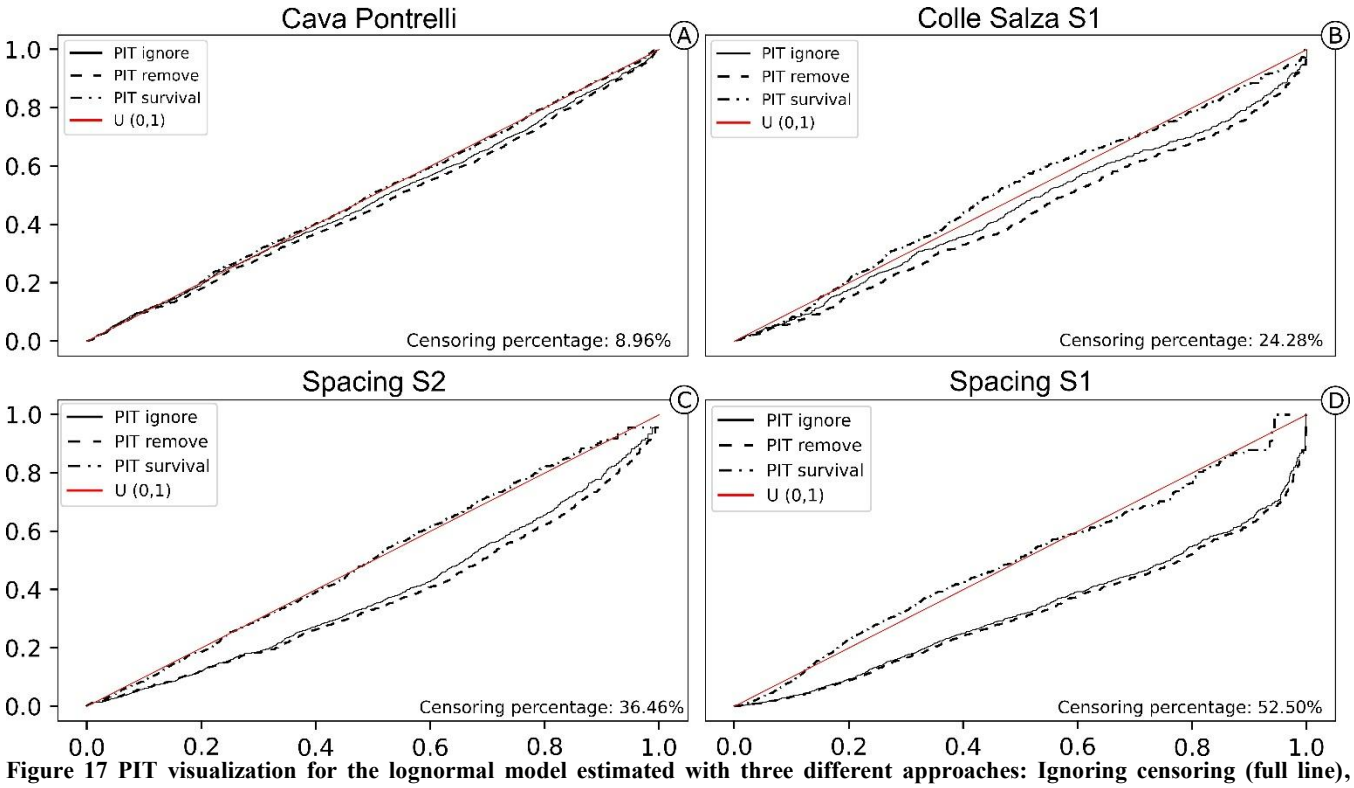


Figure 17 PIT visualization for the lognormal model estimated with three different approaches: Ignoring censoring (full line), removing censored data (dashed line), and using a survival analysis approach (dash and dot line). The survival model follows almost constantly the true model (red line) while the other two approaches increasingly diverge on longer fractures with a rate depending on the censored fraction.

Overestimation of the mean length would be possible in these scenarios when we do not consider censoring as independent from the length distribution (for example if only fractures shorter than a certain value are systematically censored). However, this would violate both the core underlying hypothesis of random censoring, and standard geological experience, and thus we do not deem it possible under these imposed limits. Furthermore, the model obtained by removing censored values is always the worst. This is because MLE is a consistent estimator that converges as the number of data increase, and thus by removing data we are effectively hindering the estimator precision (in the last case study, we would be removing 52% of the data i.e. 667 fractures). This effect is also visible when comparing the estimation of important distribution parameters such as mean (μ) and standard deviation (σ^2). Table 4 compares the values of these parameters calculated with the different methods, showing a consistent underestimation of both the mean and standard deviation, with a stronger underestimation coming from the removing censored method.

Method	**Pontrelli** (8.96%)		**Salza** S1 (24.98%)		**Spacing S2** (36.46%)		**Spacing S1** (52.50%)	
	$\boldsymbol{\mu}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\mu}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\mu}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\mu}$	$\boldsymbol{\sigma^2}$
Ignore censoring	4.077	5.032	0.9	0.895	0.844	1.255	1.373	1.453
Remove censored	3.856	4.721	0.811	0.785	0.744	0.974	1.291	1.351
Survival	4.586	6.156	1.218	1.482	1.551	2.804	3.072	3.324

Table 4 Summary table comparing mean (μ) and standard deviation (σ2) values for increasing censoring fractions. Both the ignore censoring and remove censored data methods underestimate the values with the latter always showing a stronger underestimation.

The outlined effect could be useful to predict what to expect from any survival analysis and shed some light on the exploration of the censoring mechanics that govern these types of datasets. Still, the comparison briefly discussed in this chapter is not rigorous enough to fully constrain this relationship. Even if we have a clear increase in the censoring percentage in the different datasets, the genetic factors, lithology and censoring mechanics are completely different. Moreover, the number of measures in each case study are different and thus not statistically comparable. Because of these reasons, this behavior should be isolated and modelled in controlled synthetic experiments to be carried out in the future. Nevertheless, we found that our preliminary results and the possible implications were too interesting not to be discussed.

As a third point, the readers that support and firmly believe in power law distribution of length may be shocked by the results of this paper. In all the different case studies, the power law distribution always ranked last or second to last together with the normal distribution. Power law distributions theoretically describe many natural events, and in our case if fractures for example are self-similar then the lengths could show a power law distribution (Bonnet et

al., 2001). However, it is shown that in general truncated power law models better fit natural and realistic data and thus usually only the tails of the distribution follow a power law (Clauset et al., 2009). This entails that to properly fit a power law it is necessary to estimate both the scaling parameter (α) and the minimum truncation value (x_{min}). Truncation is a constantly present bias caused by the resolution limits of the data source above or under which no data can be acquired (right and left truncation respectively). In our case, left truncation (i.e. x_{min}) is the most common type and it is caused by the pixel resolution of the images under which fractures traces cannot be seen. Due to this limitation, for a constant resolution scale, the modelled length distribution will always have an underestimation of the frequencies for the shorter lengths and thus usually only the tail of the distribution follows a power law. Adding to this bias is also the human error caused by countless different factors (one of which is the inevitable boredom of digitizing an outcrop). If the length distribution truly follows truncated power laws, then shorter fractures near the truncation limit would be much more frequent however harder to spot. Moreover, for some fracture systems, the smaller fractures are more prone to be mineral filled and potentially less obvious features on images. This usually results in fractures that are left uninterpreted even if visible and thus leading to an increase of the real x_{min} value. Truncation bias and fitting power laws is a very active field of research still with no definitive solution (Clauset et al., 2009; Deluca and Corral, 2013), thus, the implementation, testing and discussion of such solutions would lie outside this work. For the same reason a proper estimation procedure could not be included, and the results obtained in this work relative to the power law model must be taken with a grain of salt. Nonetheless, we still wanted to leave the estimation results and discussion for power laws to again show, as Healy et al. (2017) and Rizzo et al. (2017) did, how this model cannot be blindly applied to geological data without carrying out necessary important considerations.

Finally, censoring is one of many biases that influence the length measurements, and the presented work only treats this specific bias. Because of this, if other biases are present, they will be carried out also after the correction. Moreover, the underlying statistical model between different sets can be different and each fracture set has its own set of biases that influence measurements and variation in the statistics can also occur within the same set for example in proximity of a fault or local changes in rheology. Because of all these reasons, grouping all entities in the same statistical model without any kind of consideration leads to inevitable misinterpretation of the statistics and provide a wrong statistical parametrization of the whole network and thus the analysis of a fractured rock system must be carried out in an homogeneous domain, or different stationary domains must be defined before any further analysis is carried out (Bistacchi et al., 2020). By applying this divide and conquer approach, a more precise and robust characterization is assured because statistical models and inevitable biases related to different families will not mix. These observations further highlight the crucial importance of field data acquisition and geological reasoning to avoid blindly applying these methods to outcrops and make severe inferential errors.

6 Conclusions

In this paper the effects of censoring bias on the estimation of statistical length models are delineated and discussed. In particular, we demonstrated that censoring bias leads to a general underestimation of any distribution. The typical approaches found in literature are not suitable and we have shown that using survival analysis is the better way to treat censoring also in length datasets. In particular, survival analysis offers a valid alternative to the popular circular scanline method. Firstly, survival analysis methods are based on the relationship between the event of interest and the censoring mechanism. Without any underlying geometrical assumptions (basis of the circular scanline method) this methodology is quite flexible and can be easily applied in any geological case study. Secondly, the parameters that are estimated with survival analysis are directly based on the measured length (i.e. the quantity that we want to model) and are always associated with a model thus have a higher statistical significance of the scanline method. Regarding the other parametrical methods (i.e. ignoring censoring and removing censored data), we have shown that the censoring percentage heavily influences the estimation quality. In particular, with only 8.96% of censoring the two classical methods underestimate the mean and variance and the increase in censoring percentage positively correlates with such underestimation. Also, survival analysis is visibly impacted by the increase of censoring, however its output remains always stable around the natural model (i.e. the reference diagonal). Nonetheless, the influence of censoring percentage is not yet fully constrained since in this study the visible variation is also probably given by different datasets, with different genetic factors. Thus, a more robust and in-depth analysis must be carried out to fully understand this relationship. The proposed combination of PIT plots and distances calculations demonstrate an effective approach to quantitatively rank a list of length distribution models. The workflow has the objective of comparing sensible (simple) models useful to practical applications (such as DFNs) and find which one best represents the data. The theory and necessary tools to carry out a complete statistical characterization of a fracture network are included in the specifically developed FracAbility Python library that encapsulates the necessary steps to carry out as easily and quickly as possible the workflow starting from input shapefiles. The methodology and library were successfully tested in three different use cases. The first having almost no censoring bias and the last with more than 50% of the total measurements being censored. In all cases, except for the last in which one of the two spacing distributions follows a Weibull distribution, the lognormal model was the most accurate followed by either gamma, exponential or Weibull.

7 Code and data availability

The code and data are available at the GitHub repository (https://github.com/gecos-lab/FracAbility).

Statistical Assessment of the Representative Elementary Area for Fracture Areal Intensity (P21) in Digital Outcrop Models

After submitting the thesis for the review, this chapter will be submitted as a journal article

Abstract. The definition of a representative elementary volume (REV) or area (REA) of a target parameter is a fundamental step toward the upscaling of fracture network properties generated by discrete fracture network models (DFN) to an equivalent continuous medium for in situ applications in engineering geology, hydrogeology, and structural geology. The target parameter of this work is the fracture areal intensity (P_{21}), a key metric often used as a stopping criterion in stochastic DFN simulations, that is derived directly from surface data collected at natural outcrops. We propose a novel approach to define the REA as a range bounded by a lower and an upper limit. The upper limit, often overlooked but nonetheless theorized, identifies the largest representative domain, which is crucial for optimizing computational efficiency. We evaluate the REA range based on three statistical parameters, namely: the shape, mean, and variance of the P_{21} distributions obtained with progressively increasing scan area sizes. Each statistical parameter is assessed by combining formal statistical tests and diagnostic plots. Within a multi-parametric framework, the method enables a detailed analysis of the statistical behaviour of the dataset, supporting informed decisions in defining the REA range. The methodology is tested on two fractured limestone outcrops with markedly different characteristics: (i) an abandoned quarry in the Murge Plateau (Puglia, Italy) and (ii) the Lilstock Benches in the southern Bristol Channel basin.

1 Introduction

The Geometrical, mechanical and hydraulic properties of fractured rock masses are influenced by the occurrence, geometry and spatial distribution of discontinuities at every observation scale. Increasing the size of the sampling element results in a variation in almost all measured parameters of a rock mass, due to (i) the progressive inclusion of larger discontinuities or heterogeneities, and (ii) the increasing size of the sampling domain over which these parameters are averaged. Representative elementary volume (REV) or area (REA), depending on the dimensionality, represent the area/volume above which a certain parameter value becomes independent of the position and of sampling element size with which it is calculated, and thus the value can be used to constrain wider models (Bear, 1975). In other words, REV/REA mark the transition from an inhomogeneous and discontinuous behaviour of the response variable to an homogeneous and continuous one (Hudson and Harrison, 1997). Continuing to increase the size of the sampling element, theoretically speaking, another transition to a heterogeneous behaviour can be found due to the progressive inclusion of large-scale features. Both the lower and

upper limits of REA/REV are fundamental because they define the range of stability of the target parameter (Fig.1, Hudson and Harrison, 1997).

The REV/REA concept finds application in different fields of research (Huang et al., 2020; Sari, 2021). Its robust definition is crucial for numerical modelling applications involving large-scale problems (i.e. fluid reservoirs, tunnelling, mining, large rock slope instability), based on both the "equivalent continuum" approximation (Hoek and Brown, 1997; Hudson and Harrison, 1997) and the discontinuum one, where an explicit account of all fractures is not feasible and outcrop-scale rock mass properties must be incorporated in the modelled "rock matrix" (Numerical modelling of brittle fracture and step-path failure: from laboratory to rock slope scale, 2025; Agliardi et al., 2013; Gerstner et al., 2025).

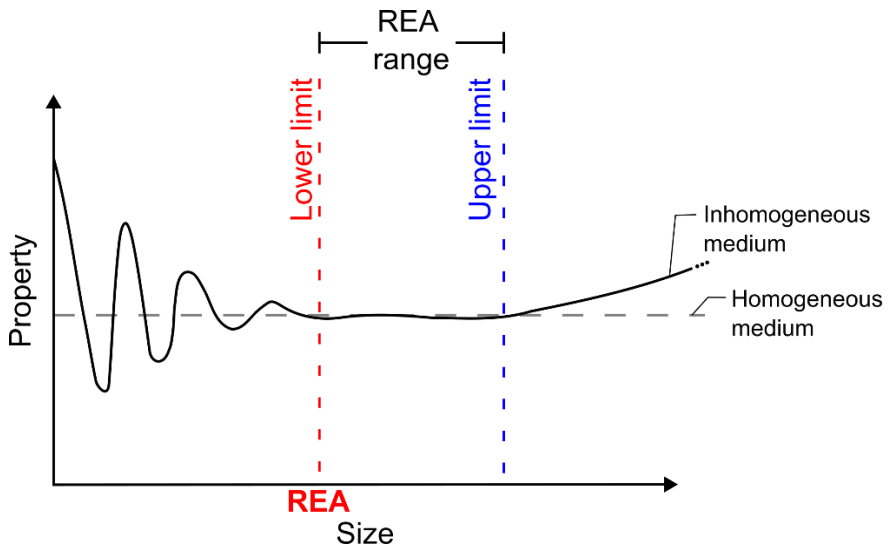


Figure 1 REV/REA theoretical behaviour, modified after Hudson and Harrison, (1997).

It has been demonstrated that there is not necessarily a unique REV for a given fracture network, but every specific parameter is associated to its own specific REV (Martinelli et al., 2020). Particularly in engineering geology literature, the REV has been determined for many parameters of interest, for example: Geological Strength Index (Marinos et al., 2005; Huang et al., 2020), damage coefficient (Ni et al., 2017), blockiness (Xia et al., 2016), fracture persistence (Song et al., 2015), among many others (Sari, 2021). In hydrogeology applications, the REV has been associated with parameters like the average block permeability (Kulatilake and Panda, 2000) and the hydraulic conductivity (Wang et al., 2002). Considering the parameters describing the geometry of fracture networks, methods for REV lower limit calculations have been proposed for the average block size (Rohrbaugh Jr. et al., 2002), mean spacing (Pariseau et al., 2008; Zeeb et al., 2013), and has been associated with the mean trace length through a scalar multiplier that goes from 3 to 8 depending on the network connectivity (Li and Zhang, 2011).

In the realm of fracture network stochastic modelling, one of the most important parameters is the fracture areal intensity P_{32}, defined as the ratio between the total fracture area and the sampling volume, or its surface equivalent - P_{21}, defined as the total fracture trace length per unit area (Dershowitz and Herda, 1992). These parameters are critical because they serve as stopping criteria in marked point process stochastic DFN (Discrete Fracture Network) modelling, meaning that fracture generation ceases once the target intensity is reached. REV calculations for P_{32} have been reported, for example, by Zhang et al., (2012) and Esmaieli et al., (2010). In such applications, REV definitions are supported by stochastic models to upscale geometrical parameters derived from outcrops into three-dimensional volumes. Alternative approaches addressing the same parameters rely directly on features mapped from outcrops (e.g., digitalized fracture traces; Martinelli et al., 2020; Casiraghi et al., 2025). In these cases, the evaluated parameter is P_{21}, with the focus placed on REA (2D) rather than REV (3D). This work focuses on the REA rather than the REV because most fracture parameters used to constrain stochastic DFN models are derived from field analogues and obtained from the analysis of outcrop surfaces (either vertical or horizontal), where fracture traces are digitized manually (Forstner and Laubach, 2022) or through automated approaches (Prabhakaran et al., 2021). Since these parameters are two-dimensional, a representative area must first be defined to ensure that parameters such as P_{21} are meaningfully estimated before being used to constrain their volumetric counterparts.

While numerous methodologies have been proposed for REV/REA calculation, they consistently focus on identifying the lower limit of the REV. However, the upper limit—beyond which the assumption of homogeneity is no longer valid—has received much less consideration. However, the upper limit of the REV is of particular importance in modelling applications because, when some property of a discontinuous rock mass is treated as a continuum and discretized. For example, in numerical methods, such as finite element or finite differences, it represents the largest cell size that can be used without losing the possibility to describe the spatial variability of this property, and hence the cell size associated with the lowest attainable computational power (the larger the cell size, the lower the computational cost). Nevertheless, a univocal definition of the upper limit of REV/REA is still elusive.

P_{21} is computed by placing scan areas of arbitrary shape (e.g., squares, hexagons, or circles) and size within the outcrop boundary, measuring the total cumulative length of fractures (or fracture segments when only partially contained), and dividing this by the area of the sampling window. By progressively increasing the scan area size (for example, the radius in the case of circular windows), it is possible to evaluate how sampling scale influences the resulting P_{21} values (Martinelli et al., 2020; Casiraghi et al., 2025). Within the REV/REA range, following the REV definition (Bear, 1975), similar statistics are expected among groups of scan areas with the same radius, and comparable statistical distributions of P_{21} value among scan areas with different radii, provided they fall within the REV range.

In this study, we present a new methodology for quantifying the Representative Elementary Area (REA) range, based on the evaluation of three statistical parameters of the P_{21} value distribution: the mean, the variance, and the distribution shape.

Each parameter is assessed using a different statistical test and an associated diagnostic plot. The shape of the distribution is evaluated using the Shapiro–Wilk test (Shapiro and Wilk, 1965), under the assumption that P_{21} values follow a normal distribution within the REA range. Test results are checked with normal probability plots. Variance equality among groups of P_{21} values obtained with different scan area sizes is tested using Levene's test (Brown and Forsythe, 1974), and results are visually inspected with residual plots (Kozak and Piepho, 2018). Differences in mean values between groups are assessed through one-way ANOVA test (Stahle and Wold, 1989).

We demonstrate the applicability of this approach on two different case studies: (i) Pontrelli quarry (Panza et al., 2019; Casiraghi et al., 2025), and (ii) the Lilstock outcrop in Bristol channel (Prabhakaran et al., 2021, and references therein). Each of these case studies exhibits distinct characteristics, enabling the method behaviour to be evaluated under different conditions and at varying scales. The Pontrelli quarry is characterized by multiple fracture sets, although only one can be clearly identified across the whole quarry pavement. The Lilstock outcrop, by contrast, is particularly clean and contains several fracture sets that together produce an extremely dense fracture network.

2 Case studies

2.1 Pontrelli quarry pavement

The Pontrelli abandoned quarry is an outcrop of fractured limestone in the Murge Plateau near Altamura (Puglia, Italy), in the forebulge of the outer Apulian platform, in front of the southern part of the Southern Apennines fold and thrust belt (Panza et al., 2019). The quarry is carved into the shallow marine intertidal limestones of the Calcare di Altamura Formation (Coniacian to Early Campanian, Panza et al., 2016). The quarry has been the subject of several studies since it offers a wide and clean pavement (around 18.000 m^2) where fractures are well exposed and the surrounding vertical walls, thus enabling the characterization of both horizontal and vertical fracture traces and the integration of information obtained from the two complementary types of exposures.

The outcrop presents three main fracture sets, connected by well-defined topological relationships (Casiraghi et al., 2025). Among the three sets, Set 1 is the most persistent and the only one detectable across the entire pavement. The fracture traces of Set 2 and Set 3 on the pavement are partly hidden by noise from quarrying activities, although they can be observed on the adjacent wall. This effect cannot be resolved simply by isolating the no-data zones,

since in areas where Sets 2 and 3 are not identifiable. Set 1, i.e. the most persistent and clearly visible, as explained in Casiraghi et al. (2025), is the subject of our test (Fig. 2).

2.2 Lilstock outcrop – Bristol channel

The Lilstock outcrop is located in the southern coast of the Bristol Channel in West Somerset (UK), near the hamlet of Lilstock (Prabhakaran et al., 2021, and references therein). The outcrop is composed of weakly deformed Jurassic sediments, and it is characterized by large-scale folds, faults, veins and fractures (Passchier et al., 2021). The fracture network consists of eight distinct fracture sets (Passchier et al., 2021), and the exceptional cleanliness of the pavement allows for the mapping of every single fracture, with virtually no gaps in data zones disrupting the network (Fig. 2). The analysis is performed on the Area E of the Bench IV (Passchier et al., 2021), a portion of the outcrop with a size comparable to the part of Pontrelli quarry pavement free of no-data zones. The fracture trace shapefile is provided as a supplementary material by Prabhakaran et al., (2021), and obtained with an automatic digitalization method. The interpretation boundary, not provided in the dataset, is instead digitalized as a "convex hull", joining the fracture trace tip lines. This would be an issue if the goal of the analysis was, for instance, to determine the fracture length distribution, as the assumption of independence between the censoring mechanism and the phenomenon would no longer hold (Benedetti et al., 2025). In the case of REA estimation, however, the interpretation boundary simply delineates the area where data are observable from the area where they are not.

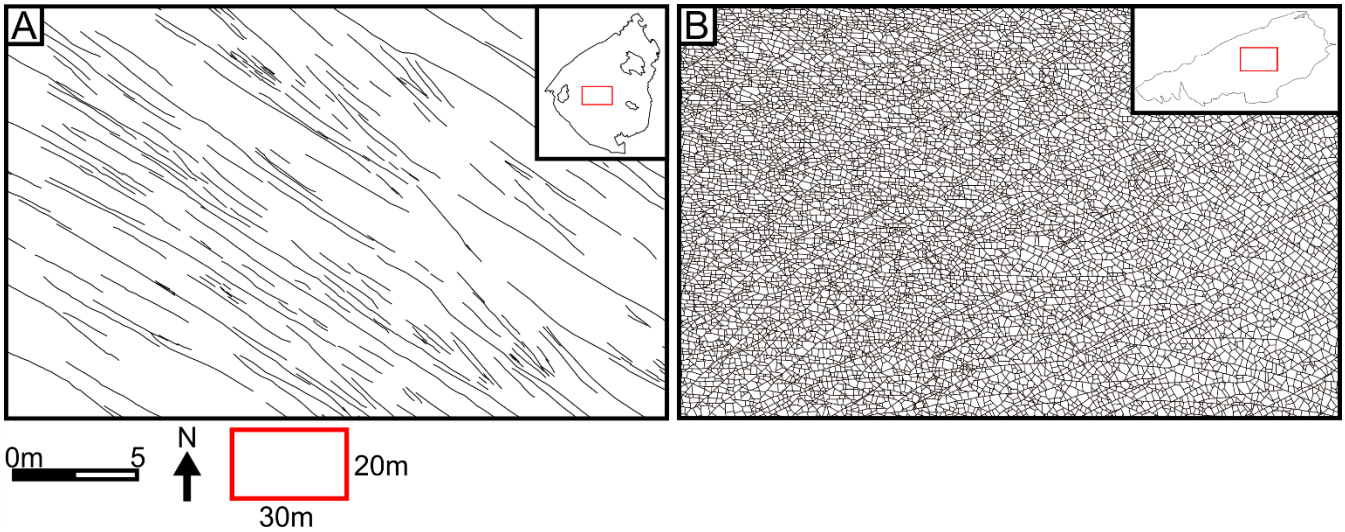


Figure 2 Close-up view of the digitized fracture network for the Pontrelli quarry (A) and the Lilstock outcrop (B). Only the fracture sets considered in the analysis are shown: Set 1 for the Pontrelli quarry and all eight recognized fracture sets for the Lilstock outcrop (Passchier et al., 2021).

3 Methods

Analysis of Variance (ANOVA), first introduced by Ronald Fisher, is a well-established statistical methodology that finds applications in a wide range of research fields and experimental design (Stahle and Wold, 1989; Mooi et al., 2018; Rajaraman et al., 2019). It aims to evaluate the effect of different factors on a response variable. Aligning this concept to our problem, the response variable is P_{21}, while the factor under investigation is the size of the scan area radius. Therefore, the number of factors in the proposed experimental design will be equal to the number of scan area radii needed collect P_{21} values from the outcrop.

The upper limit of scan area radius range is defined by the outcrop size (the scan area cannot be larger than the outcrop itself), while the lower limit is qualitatively defined by checking preliminary results, for example with the Tukey (1977) box plot method. Decreasing too much the scan area radius when the number of outliers is already significant will result in a waste of computational power. Having defined the range, scan area radii are here defined through equal-spacing sampling between the maximum and minimum values (but different spacing could be used in theory).

ANOVA requires different underlying assumptions to be validated. The basic assumptions that allow working in the ANOVA framework are: (i) the continuity of the response variable and (ii) the independence of each measurement. The continuity assumption is supported by the nature of P_{21}, which is a numerical parameter ranging from 0 to infinity and can theoretically be measured at infinitely many values within this range. The independence of observations within each factor is ensured by employing a random sampling with replacement strategy. This approach guarantees that the selection of one sampling unit does not influence another, thereby satisfying the statistical requirement for independence regardless of the underlying spatial correlations of the fracture network. In this approach, scan area centres are randomly positioned within the interpretation boundary. If a scan area intersects the interpretation boundary, it is rejected and randomly repositioned until the conditions are met, that are: (i) complete inclusion into the interpretation boundary, or (ii) tangency, that is permitted to ensure equal sampling probability across the outcrop. This step represents the most computationally demanding part of the workflow. Increasing the scan area radius leads to more frequent rejections due to boundary intersections, and scan areas larger than the interpretation boundary are systematically rejected. This process, and the following testing procedures are repeated one hundred times to account for the variability of the random sampling.

With regard to grid sampling strategy, for example that employed in Casiraghi et al. (2025), or Martinelli et al., (2020) random sampling allows the same number of P_{21} data to be collected for every scan area radius. Even if it is not an underlying assumption, equal sample size is a desirable condition when it comes to testing different groups of data. Sample size is related to the power of a statistical test and comparing the test results of samples with different sample sizes will produce misleading results (Kozak and Piepho, 2018).

Aside from continuity and independence, that are assumptions related to the nature of the response variable (continuous, discrete, categorical etc.) and sampling strategy (e.g. random sampling with replacement, grid sampling), ANOVA test requires two additional assumptions that are: (i) the normality of the error factor in the ANOVA linear model (Eq. [4]), and (ii) the homogeneity of variances between the tested factors. In other words, it is first necessary to assess whether the P_{21} values collected using a given scan area radius (e.g., 1 m; Fig. 3) can be reasonably described by a normal distribution, and subsequently to evaluate whether datasets obtained at different radii (e.g., 1 m, 2 m, 3 m) exhibit comparable variances.

In order to test these two assumptions (Stahle and Wold, 1989), we propose integrating formal significance tests with diagnostic plots. Normality assumption is tested with the Shapiro-Wilk (Shapiro and Wilk, 1965) test and normal probability plots (Kozak and Piepho, 2018), while the homogeneity of variances is tested with the Levene test (Brown and Forsythe, 1974) and residual plots (Kozak and Piepho, 2018). The Shapiro-Wilk test, Levene test, ANOVA test and normal probability plots are performed using the SciPy stats package (https://docs.scipy.org/doc/scipy/reference/stats.html).

3.1 Data collection

The datasets are composed by an interpretation boundary (polygon shapefile) and a series of fracture traces (polyline shapefile). P_{21} is calculated on digitalized fracture traces from TS-DOMs (Textured Surface Digital Outcrop Models, Casiraghi et al., 2025). P_{21} values are straightforwardly obtained by calculating the total sum of length of the fracture traces (complete or partial) inside the circular scan area and dividing it by the scan area surface area. P_{21} is not calculated using estimators (e.g. Mauldon et al., 2001). For any fixed scan area radius, we collect a random sample of size n of such areas and compute the variable P_{21} (Fig. 3). The number of scan area radii depends on the interpretation boundary size and the radii discretization strategy, that means the number of radii collected between the maximum and minimum radius, and the stepping strategy (e.g. equally spaced). The procedure is repeated m times, each time with a new and independent random sample. Consequently, the statistical tests, which will be described in the following sections, are applied m times. In the present simulation $n = m = 100$. This kind of "meta-analysis" allows for consideration of the variability inherent in random sampling while ensuring robustness to the statistical analysis.

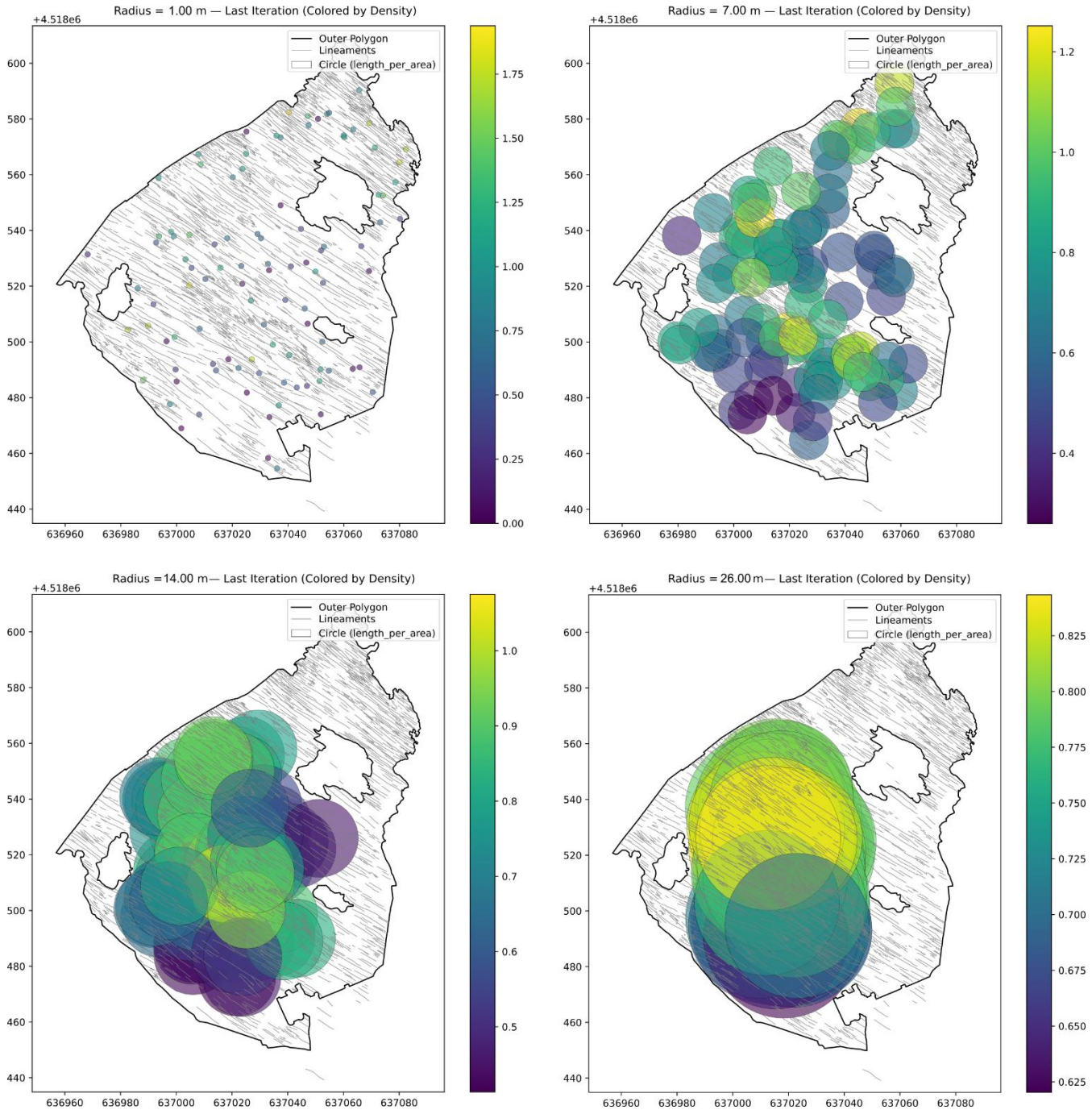


Figure 3 Example of P_{21} data collection at progressively increasing scan area radius for the Pontrelli quarry outcrop. One hundred scan area are randomly distributed inside the interpretation boundary for every scan area radius. At small scan area radii, the scan areas can be placed even in the most convex part of the outcrop. At increasing radii, the position of the scan area is progressively constrained by the shape of the interpretation boundary and by the presence of no-data zones inside the outcrop. This process represents one realization.

3.2 Exploratory data analysis

Before performing the testing procedure, data collected from one or more realizations (if necessary) are first analysed using graphical tools to identify potential outliers or anomalous patterns. This preliminary step also provides an initial understanding of how P_{21} evolves with increasing scan area radius, thereby supporting a more critical assessment of the statistical tests and diagnostic plots presented in the following sections. The graphical tools employed include: (i) variance plots, showing the relationship between sampling variance and scan area radius, (ii) mean plot showing the mean variation at increasing scan area radius, and (iii) boxplots of P_{21} values for each scan area radius, constructed following the method of Tukey (1977).

3.3 ANOVA mathematical model

The ANOVA test aims at assessing the effect of different factors on the response variable behavior (Stahle and Wold, 1989). In other words, the ANOVA test is aimed at detecting significant differences in mean between groups of data. The easiest configuration is therefore a situation in which every factor has the same mean and every observed difference is due to random errors. Given a certain number of factors the model would be:

$$x_{i,j} = \mu + e_{ij} \quad [1]$$

Where $x_{i,j}$ is the observation i belonging to factor j, μ is the mean of the considered variable populations and e_{ij} is the random error associated with observation i, belonging to factor j. Adding a level of complexity, a configuration in which differences are actually related to the effect of different factors:

$$x_{i,j} = \mu_j + e_{ij} \quad [2]$$

Where μ_j is the mean of the group of observations related to factor j. It is possible to explicitly express the effect of factor j as:

$$\alpha_j = \mu_j - \mu \quad [3]$$

Where α_j is the effect of the factor j on the response variable. Adding [2] into [3], it yields:

$$x_{i,j} = \mu + \alpha_j + e_{ij} \quad [4]$$

Equation [4] is the so-called ANOVA linear model, an additive model that consist of the sum of three effects: (i) μ, the common effect, (ii) α_j, the effect of the j-th factor, and (iii) e_{ij} the random error associated to each observation of each factor (Stahle and Wold, 1989).

3.4 Normality assumption

The normality assumption of the ANOVA test involves the error term of the linear model (e_{ij}), described in equation [4] (Kozak and Piepho, 2018). The theoretical distribution of the errors is unknown, but errors can be estimated with the residuals, defined as:

$$\hat{\varepsilon}_{i,j} = y_{i,j} - \mu_j \quad [5]$$

Where $\hat{\varepsilon}_{i,j}$ is the residual associated with the *i*-th observation of the *j*-th factor, $y_{i,j}$ is the *i*-th observation of the *j*-th factor and μ_j is the mean of the group of observations related to factor *j*, as already explicated in equation [3].

Normality assumption is tested with the Shapiro-Wilk test (Shapiro and Wilk, 1965). The test is based on the W statistic that measures the goodness-of-fit to a normal distribution. The values of this statistic range from 0 to 1.0, where values close to 1.0 indicate a normal distribution of the residuals. Significance level for the test is set at 0.05.

Associated with the statistical test, normal probability plots for every tested factor are also provided. Normal probability plots are a graphical procedure to check if data are approximately normally distributed, and they are a specific case of Q-Q plots, where the ordered sample percentiles are plotted against the theoretical percentiles of a normal distribution. If data are normally distributed, points should approximately plot along a straight line. From the shape of the distribution in a normal probability plot, it is possible to detect other characteristics of the dataset, for example the presence of outliers, left or right skewed residuals, or heavy-tailed residuals (presence of both extreme positive and negative values). According to the expected qualitative behaviour of the REA, at small scan area radii the P_{21} distribution is expected to be characterized by extreme values, with a significant amount of zeros on one side and high P_{21} values on the other. This asymmetrical configuration typically corresponds to a right-

skewed distribution. Within the REA range, where the heterogeneous medium can be approximated as homogeneous, a more symmetrical distribution is expected, with P_{21} values distributed approximately around the mean. At the upper limit of the REA range, the progressive inclusion of large-scale features leads to a loss of symmetry, producing a different data distribution (it varies on a case-by-case basis), in which zeros are not to be expected.

3.5 Homogeneity of variances assumption, REA upper limit definition

Homogeneity of variances or homoscedasticity is a condition in which the variances of the different factors are comparable. Homoscedasticity is tested with the Levene test (Brown and Forsythe, 1974), performed on residuals (Eq. [5]). While Shapiro-Wilk test is straightforwardly applied to every factor, application of Levene test is controlled by two variables. On one hand there is the number of factors to be tested. The minimum number of factors is equal to three, that is also the minimum number of groups that can be tested with the ANOVA test. The upper limit is not defined, but qualitatively it is possible to say that the higher the number of tested factors, the higher the probability of finding one group that is significantly different from the others. On the other hand, there is the "location". Considering a sequence of scan area radii of increasing size, the Levene test can be conceptualized as a moving window, applied progressively from smaller to larger scan areas. According to REA theoretical behaviour, small scan areas exhibit high variability due to the heterogeneous behaviour of the response variable, leading to rapidly changing group variances and frequent test rejections. This is followed by a stability range (the REA range), where the variance across factors becomes qualitatively stable, resulting in an increased rate of test acceptance. Beyond the upper limit of the REA range, acceptance rates are expected to decrease again as large-scale features introduce additional variability. To capture this theoretical behaviour, Levene test is applied to every consequent group of scan area radius, in an ordered way, from the smallest windows to the largest ones. Group size is iteratively increased starting from groups of three factors, until acceptances are reduced to zero.

Levene test results are complemented by residual plots, which serve as diagnostic tools to assess differences in residual variance across factors. In this framework, standardized residuals—calculated by dividing each residual by the standard deviation of the group of residuals—are plotted against factors (scan area radii), treated here as categorical variables. Standardization centres the residuals around zero, facilitating comparison among groups. Homoscedasticity is assumed when the random scatter around the zero line is qualitatively similar across groups. Following Kozak and Piepho, (2018), a highlighted band corresponding to ±1.96 of the estimated standard deviation of the standardized residuals is added to aid visual inspection. Funnel- or cone-shaped patterns in the plots indicate substantial variance heterogeneity (Kozak and Piepho, 2018).

3.6 Testing procedure and ANOVA test

Having described the core tools of the presented statistical method, it is now possible to describe the testing procedure for REA range definition. The normality of the n P_{21} values collected for any fixed scan area radius is tested with the Shapiro-Wilk test. Acceptances and rejections, for the m realizations, are collected and represented by a bar (see Fig. 7). This means that for any fixed scan area radius we have a bar representing the outcome of the m test: orange bars represent acceptance, grey ones represent rejection. The normality assumption is considered plausible when the acceptance ratio is around 80%. The normality of P_{21} of these scan area radii, that meets the previous condition is also checked with the probability plots. Similarly, the Levene test is iteratively applied from the smallest radius to the largest, progressively increasing the number of tested radii. The number of stacked bar plots for the Levene test is equal to the number of times the group size has increased. The largest group of factors maintaining a significant level of acceptance will be defined as the candidate REA range to be tested with ANOVA. In the same way as for the normality assumption, Levene test results are also checked with the residual plots. The accepted group is cross-validated using the Shapiro–Wilk test results, as each factor accepted by the Levene test must also satisfy the normality assumption. If any factor within the group fails to meet the normality criterion, the group is progressively reduced until a group of consecutive factors fulfils both the homoscedasticity and normality assumptions.

The ANOVA test is applied to the accepted group of data. The ANOVA core test is also called the F-test, which is built such that the value of the F-statistics is large if significant differences are detected between the tested factors (Stahle and Wold, 1989). If the ANOVA test is accepted with a significance level of 0.05, then it is possible to assume the tested groups of factors as the range of stability of the P_{21} parameter, i.e. the REA range.

4 Results

4.1 Pontrelli quarry

The Pontrelli quarry pavement extends approximately 120 m NW–SE and 150 m NE–SW. In principle, this would allow for the use of relatively large scan areas, with an upper radius limit of about 60 m. However, the presence of no-data zones within the pavement restricts the maximum scan area radius to ~26 m (Fig. 3). The lower limit is set to 1 m, which, as shown by the boxplot results (Fig. 4), is sufficiently small to yield both a considerable number of empty scan areas and outliers. Based on the boxplots, the variance plot (Fig. 5) and the mean (Fig. 6), it is possible to qualitatively define three zones. The first zone is characterized by strong variance and mean fluctuations, and asymmetrical distributions. This zone extends to a radius of approximately 7 m. This is followed by a zone of relative stability, that goes from 7m in radius to 13-15m (depending on the iteration). In this zone the mean P_{21} stabilizes around 0.71-0.73 m^{-1}. Finally, as the scan area radius approaches the maximum size allowed by the

interpretation boundary, scan areas can no longer be freely positioned across the outcrop (Fig. 3), and the P_{21} distribution collapses toward a single value, corresponding to the only feasible position. In this final stage, variance tends toward zero and most P_{21} values are classified as outliers (Fig. 4).

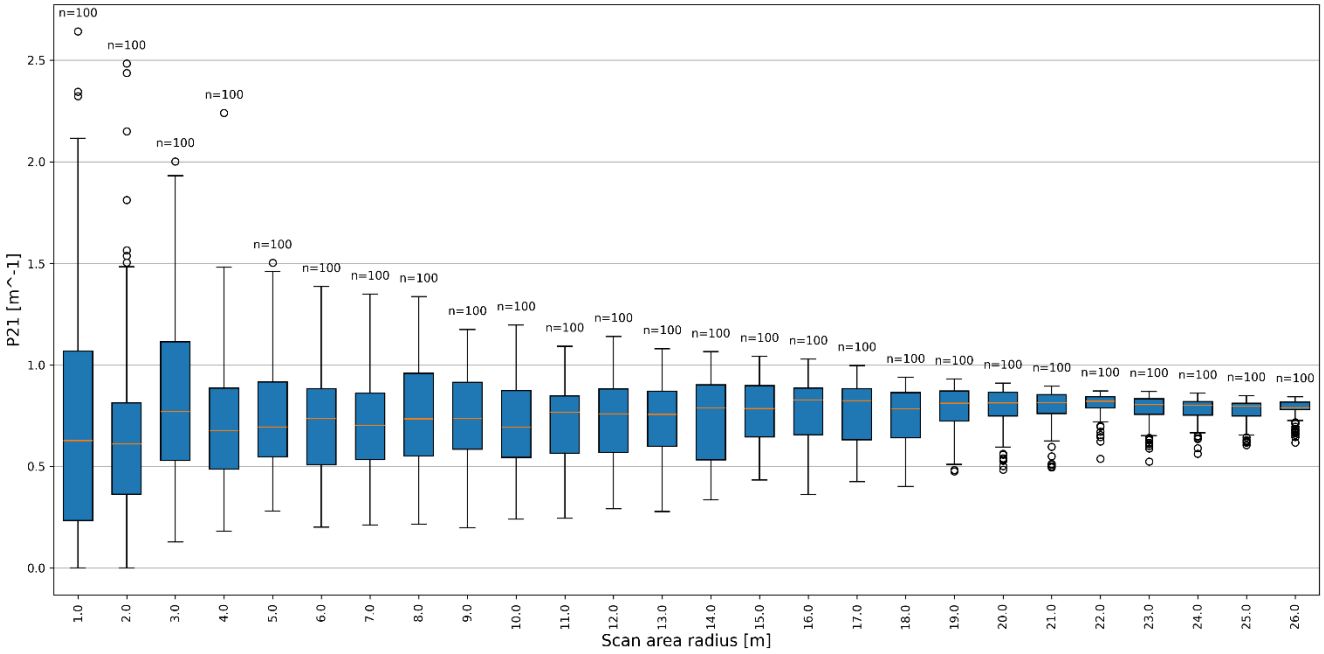


Figure 4 Boxplot of P_{21} data collected at increasing scan area radius, for Pontrelli quarry outcrop.

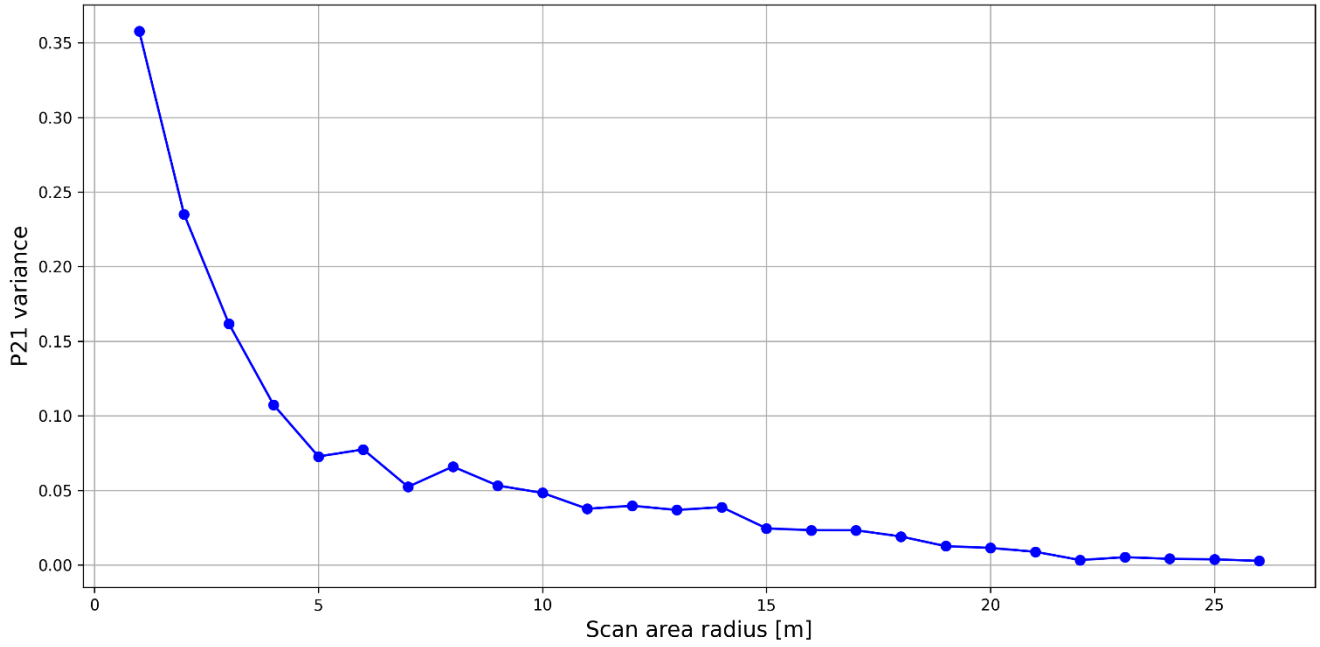


Figure 5 Variance plot of P_{21} statistical distribution for Pontrelli quarry outcrop.

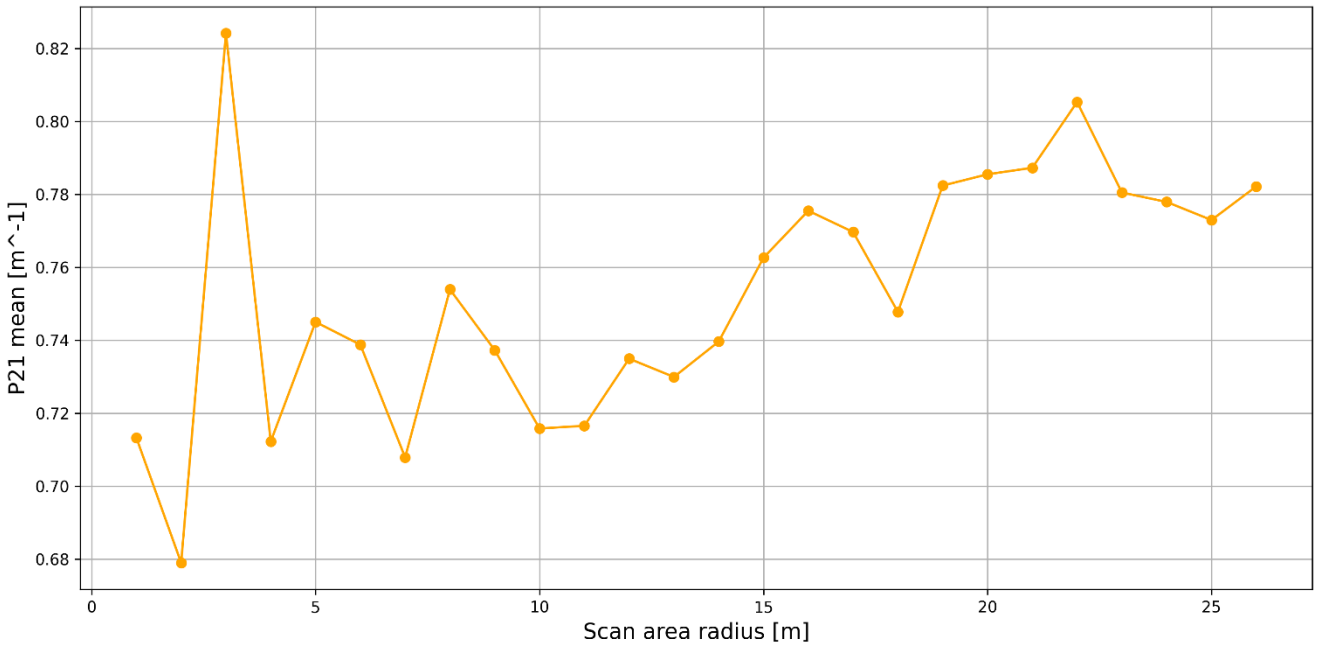


Figure 6 Mean plot of P_{21} statistical distribution for Pontrelli quarry outcrop.

The Shapiro-Wilk test results show that the best performing scan area radius are 8m and 9m with acceptances above the 80% threshold (Fig. 7). 7m and 10m are few acceptances away from the threshold, and at 6m and 11m a huge drop in acceptances is detected, falling below 50%. This result well matches the qualitative assessment made with the boxplots, the variance and the mean plot (Fig. 4, 5, 6). Scan area radius from 7m to 10m are checked with the respective probability plots (Fig. 8A, B). From the probability plots strong deviation from the normal model (e.g. heavy tails, left or right skewed distributions) cannot be detected and therefore the normality assumption is confirmed.

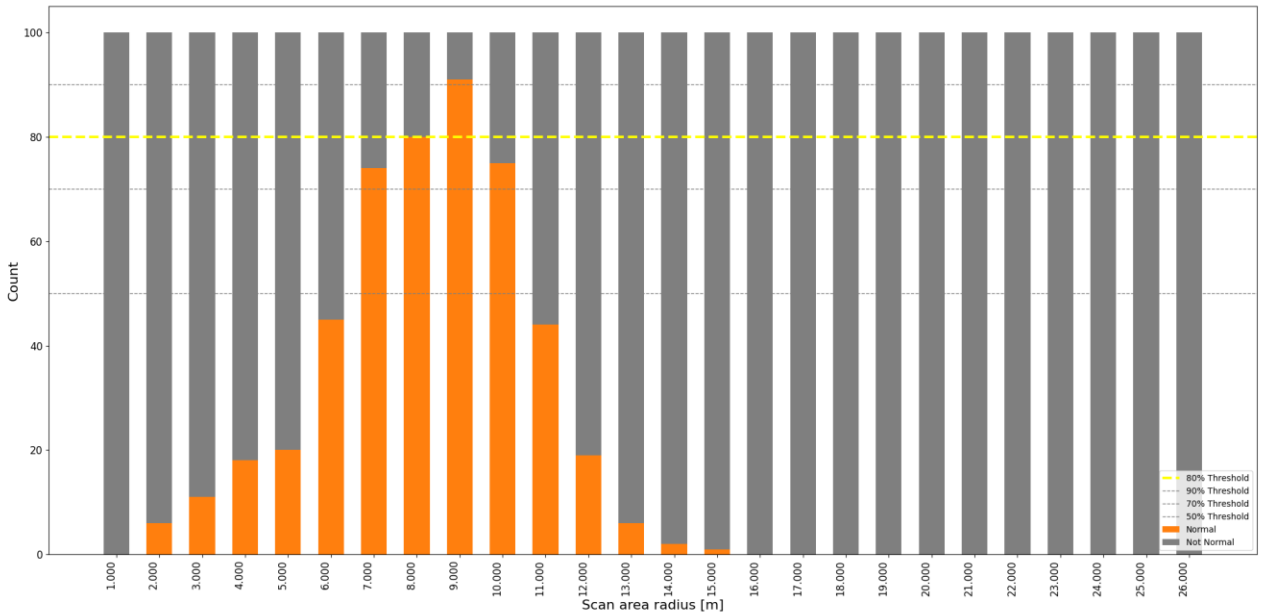


Figure 7 Shapiro-Wilk test results for Pontrelli quarry outcrop. The orange bars represent acceptances, while the corresponding grey bars indicate rejections. Their sum always equals 100, corresponding to m, the total number of realizations .

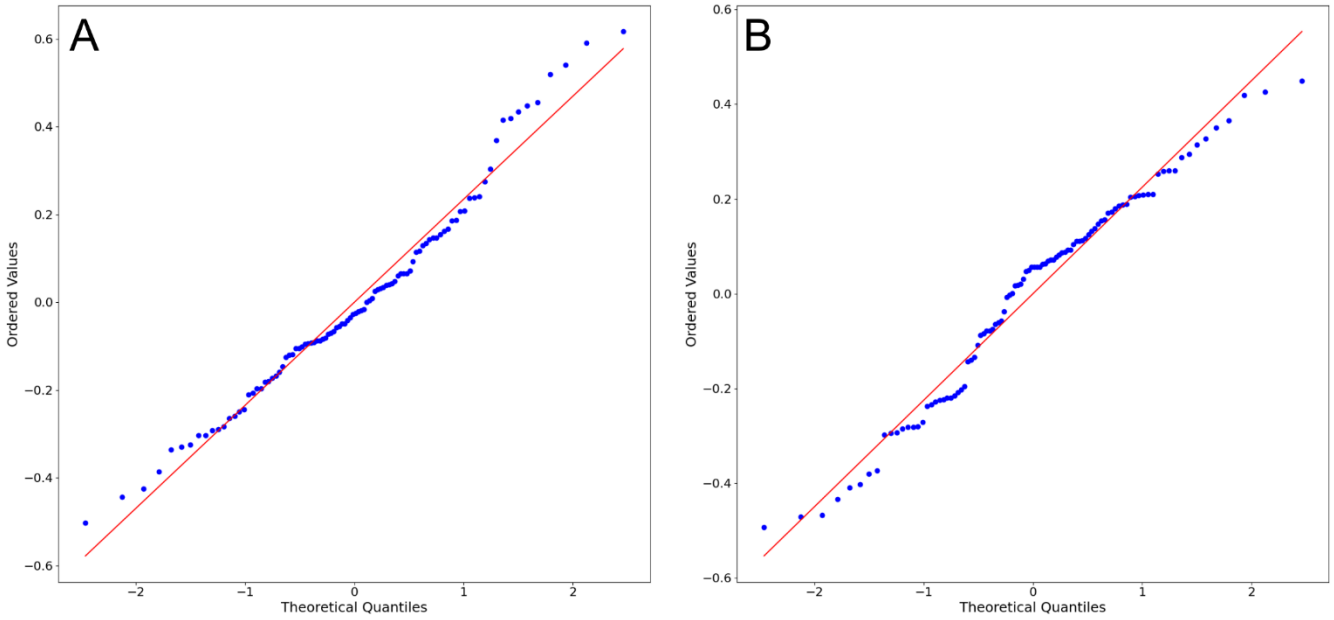


Figure 8 Normal probability plots for scan area radius equal to 7 m (A) and 10 m (B) in Pontrelli quarry outcrop.

Levene test results show that the tested groups remain representative up to a range of five factors (Fig. 9C). From six factors onward acceptances decrease by a significant amount (Fig. 9D). Between the size-five groups, the candidates that keep an acceptance ratio around 80% threshold are: 7m to 11m (close to 80% threshold), 8m to 12m,

and 9m to 13m. Homoscedasticity is supported by the residual plot, as each accepted group displays a region with almost straight boundaries (Fig. 10).

By cross-validating these candidate groups with the Shapiro-Wilk test results, it is possible to see that each group contains at least one factor that does not match the normality assumption (11m for the first group, 11m and 12m for the second group and 11m 12m and 13m for the last group). In the size-four groups (Fig. 9B), the 7m to 10m group both keep an acceptance ratio above the 80% threshold, and every factor is accepted for normality.

Figure 9 Levene test results for Pontrelli quarry Outcrop for different group sizes. (A) Group size = 3, (B) Group size = 4, (C) Group size = 5, (D) Group size = 6.

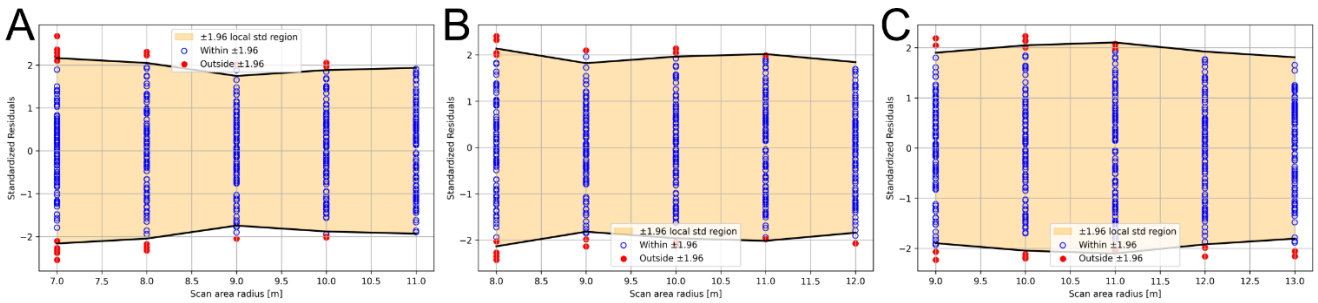


Figure 10 Residual plots for Pontrelli quarry outcrop. (A) Scan area radii from 7 m to 11 m, (B) Scan area radii from 8 m to 12 m, (C) Scan area radii from 9 m to 13 m.

The ANOVA test on the 7m to 10m group is accepted with an acceptance ratio way above 80% (Tab. 1). Finally, it is possible to conclude that, given the convergence between the Shapiro-Wilk, Levene and ANOVA test results, and the validation through the diagnostic plots and exploratory data analysis, the REA range for the Pontrelli quarry pavement is between scan areas that go from 7m to 10m in radius.

	ANOVA acceptances	ANOVA rejections	Test result
7m to 10m group	95	5	accepted

Table 1 ANOVA test result for Pontrelli quarry Outcrop.

4.2 Lilstock outcrop - Bristol channel

The Lilstock outcrop in Bristol channel is characterized by a surface area of the same order of magnitude as the Pontrelli quarry, but it presents a much denser fracture network, composed of multiple fracture sets. As an initial try, the lower limit of the sampling window radius is set to 0.005m (qualitatively equal to the average total spacing), while the upper limit is set to 18m. Given the higher fracture density, we densified the number of collected scan area radius to 50, to try to catch every change in the P_{21} distribution behavior.

Exploratory data analysis shows that, after an initial sharp decrease between the first and second scan area radii, the mean is stable around 8.5 $m^{-1} \pm 0.25$ m^{-1} (Fig. 11, 13). The sharp drop is also captured in the variance plot (Fig. 11, 12), after which the variance decreases steadily with a gentle slope. This trend is likely related to the progressive increase in scan area radius.

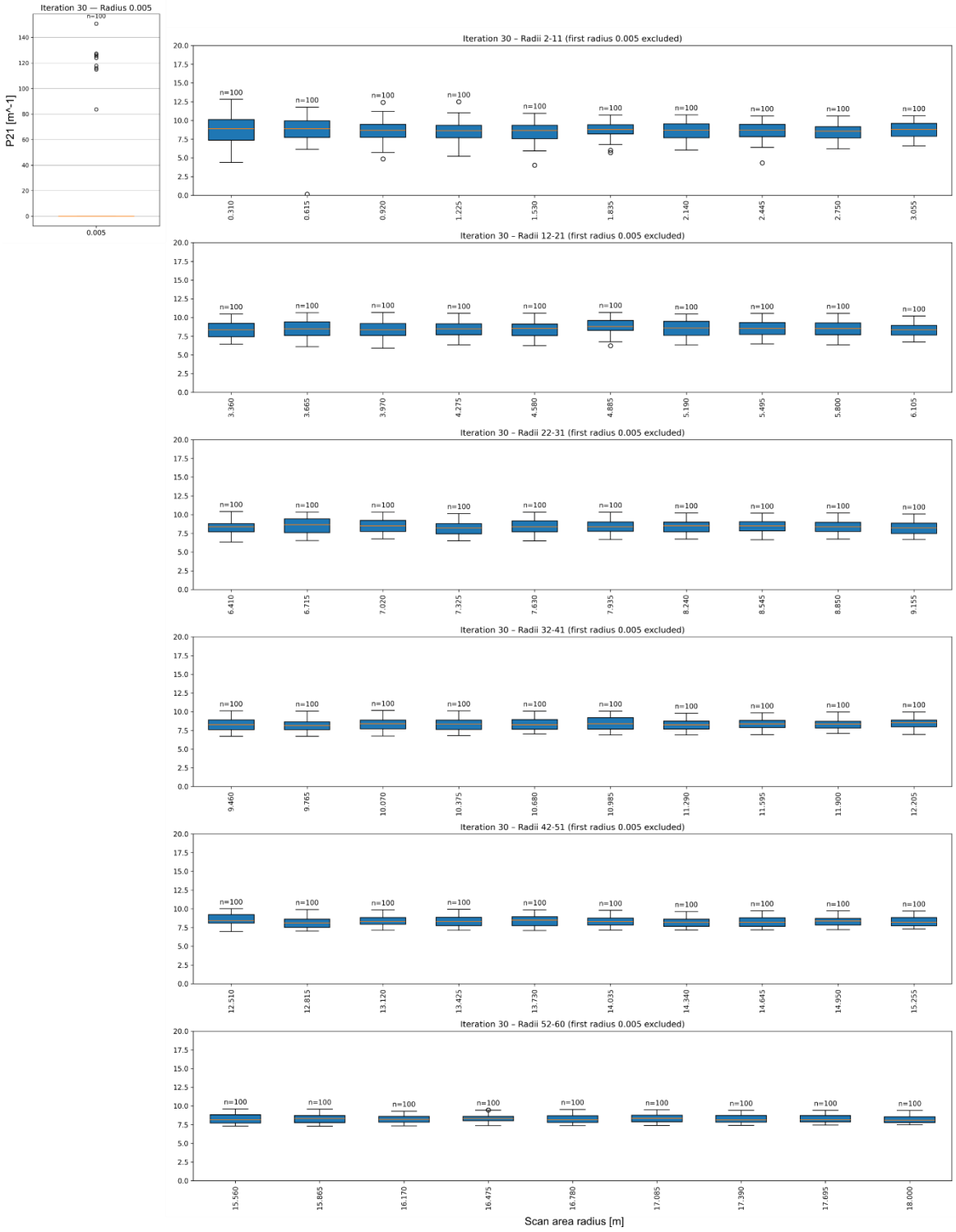


Figure 11 Boxplot of P_{21} data collected at increasing scan area radius, for Lilstock outcrop. The first boxplot is shown separately so that its larger scale does not compress the remaining boxplots and hinder readability.

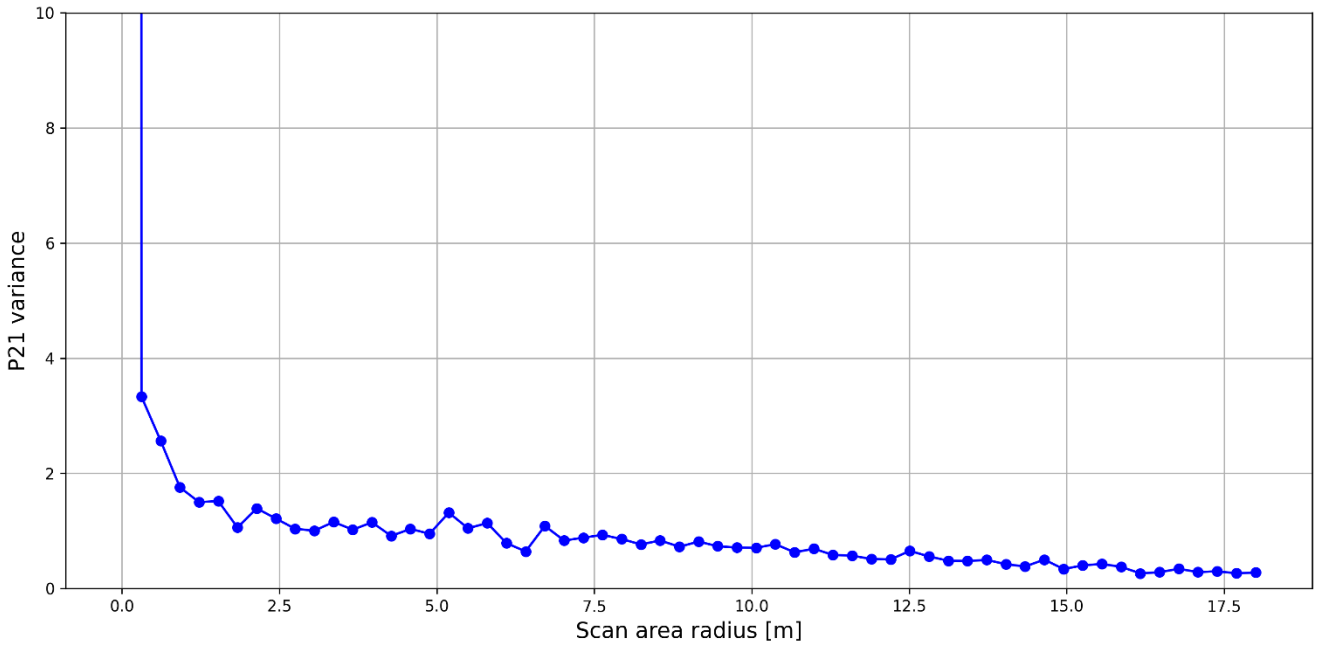


Figure 12 Variance plot of P_{21} statistical distribution for Bristol Channel outcrop. The y axis is limited to y = 5, to better highlight the variance trend. For scan area radius equal to 0.005 m the variance is equal to ≈1200.

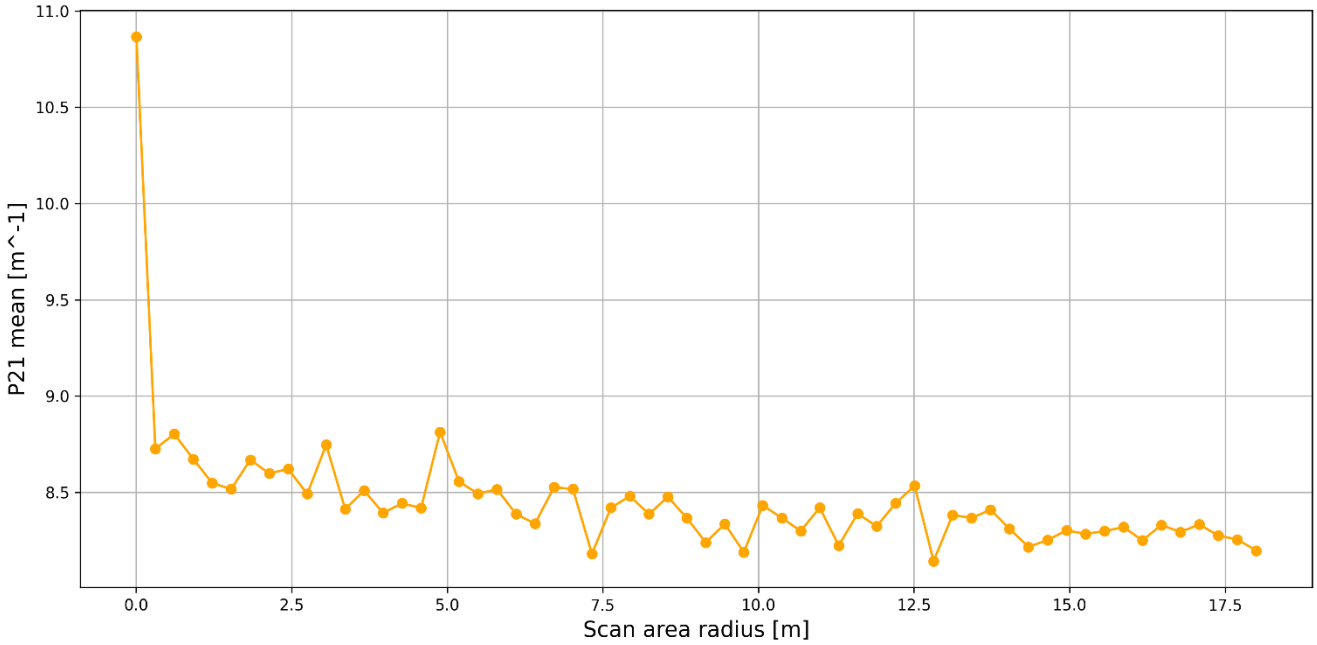


Figure 13 Mean plot of P_{21} statistical distribution for Pontrelli quarry outcrop.

Normality is never achieved, as indicated by the Shapiro–Wilk test results, with the highest acceptance ratio (around 70%) corresponding to the second smallest scan area radius (Fig. 14). Inspection of the normal probability plots reveals that deviations from normality arise from local overestimation or underestimation zones, or from areas

where P_{21} values are clustered or align along an almost straight line, suggesting that the scan areas are sampling nearly identical P_{21} values (Fig. 15). No strong departures from normality, such as pronounced left- or right-tailed distributions, are observed as, for instance, in the largest scan area radius at Pontrelli quarry. As results by inspecting the normal probability plots, although the P_{21} distributions are not perfectly normal, they are approximately symmetrical independently on the scale of observation (Fig. 15).

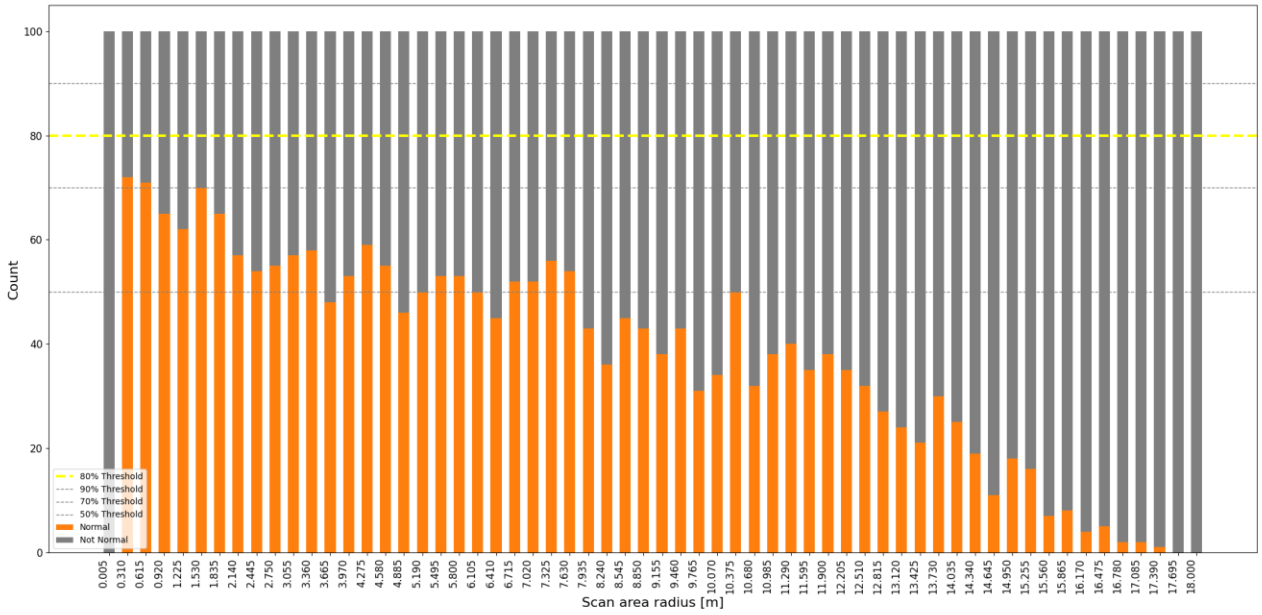


Figure 14 Shapiro-Wilk test results for Bristol channel outcrop.

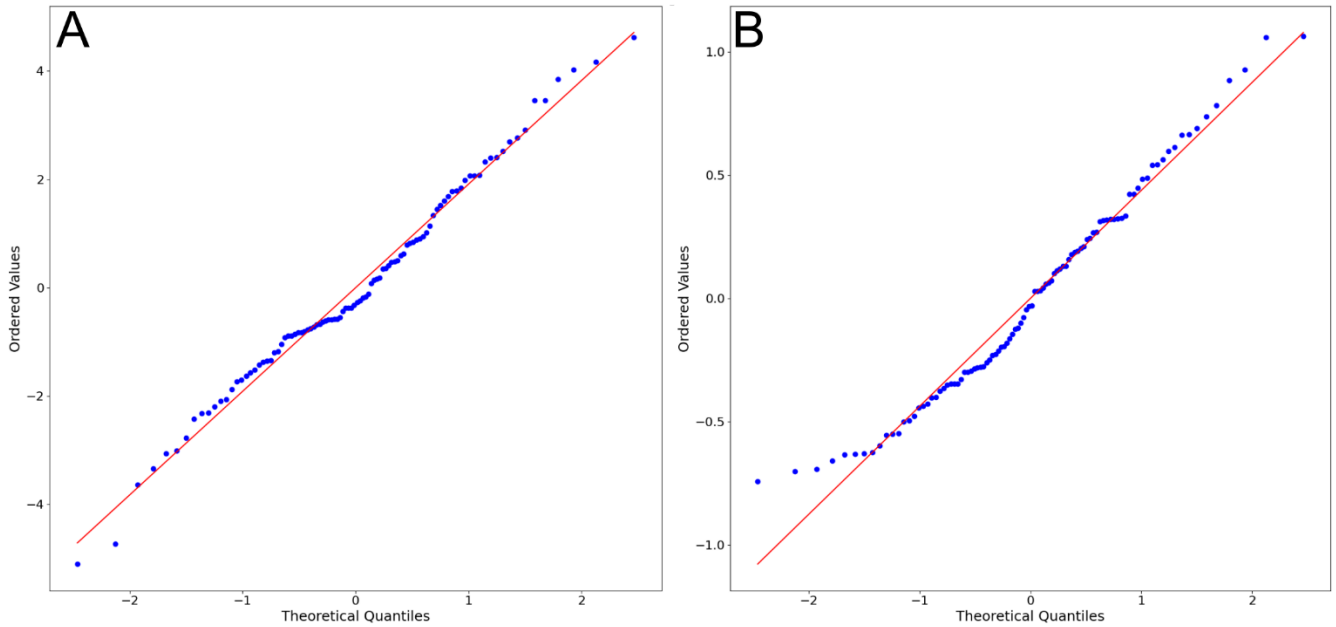


Figure 15 Normal probability plots for scan area radius equal to 0.31 m (A) and 18 m (B) in Bristol channel outcrop.

Levene test results show that homoscedasticity maintains a significant level of acceptance up to size-fourteen groups, after which acceptance rates progressively decrease from size-fifteen groups onward, likely reflecting the variance reduction observed in the variance plot (Fig.16A, B). This high degree of acceptance contrasts with the behaviour observed at the Pontrelli quarry, where acceptance rates sharply decline beyond the REA range (Fig. 9).

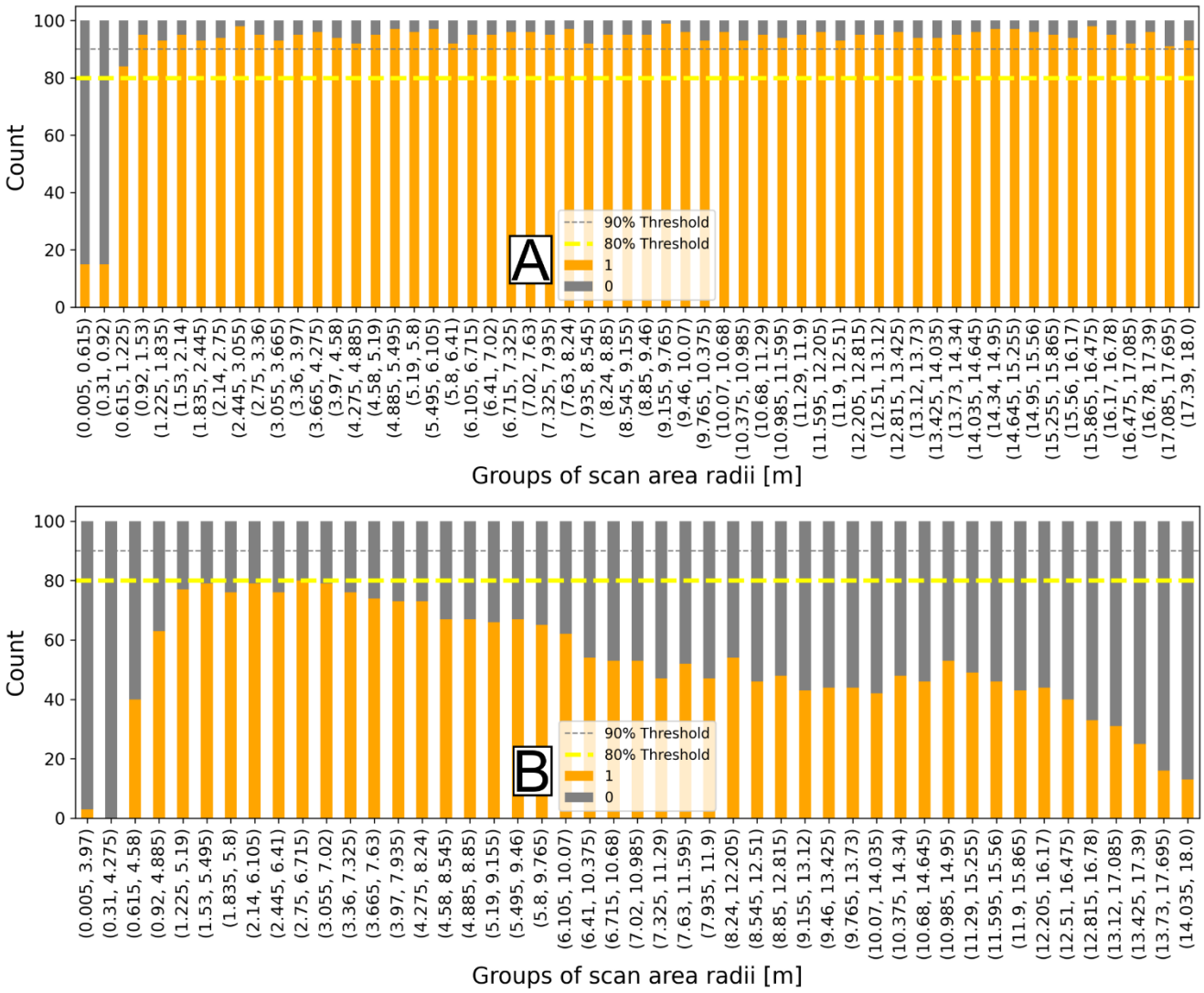


Figure 16 Levene test results for Bristol channel outcrop for different group sizes. (A) Group size = 3, (B) Group size = 14.

The fracture network in the Bristol channel outcrop therefore exhibits a markedly different behaviour from set 1 fractures in Pontrelli quarry. Here, the P_{21} distributions are not normal but remain symmetrical, with a stable mean P_{21} and variances that stay approximately equal even across large group sizes. It is likely that, if the influence of scan area sizes approaching the boundary limit could be removed, high acceptance levels would persist for even larger groups. The distributions across varying scan area radii display consistent self-similarity in terms of both

shape and variance, supporting the hypothesis of a P_{21} distribution that remains homogeneous irrespective of the scan area size.

What remains to be examined is the unusual behaviour observed in the mean and variance plots and in the statistical tests between the first (0.005 m) and second (0.31 m) scan area radii. In this range, there is a sharp drop in both the P_{21} mean (Fig. 13) and variance (Fig. 12). The Shapiro–Wilk test shows an increase from 0% acceptances at 0.005 m to 70% at 0.31 m (Fig. 14), while the Levene test rises from nearly 0% to over 90% acceptances for size-three groups (Fig. 16).

To investigate this transition in detail, the list of scan area radii was expanded by sampling ten intermediate radii between 0.005 m and 0.31 m. This new range of scan area radii reveals an acceptance pattern very similar to that observed at the Pontrelli quarry, but at a much smaller scale (Fig. 17,18).

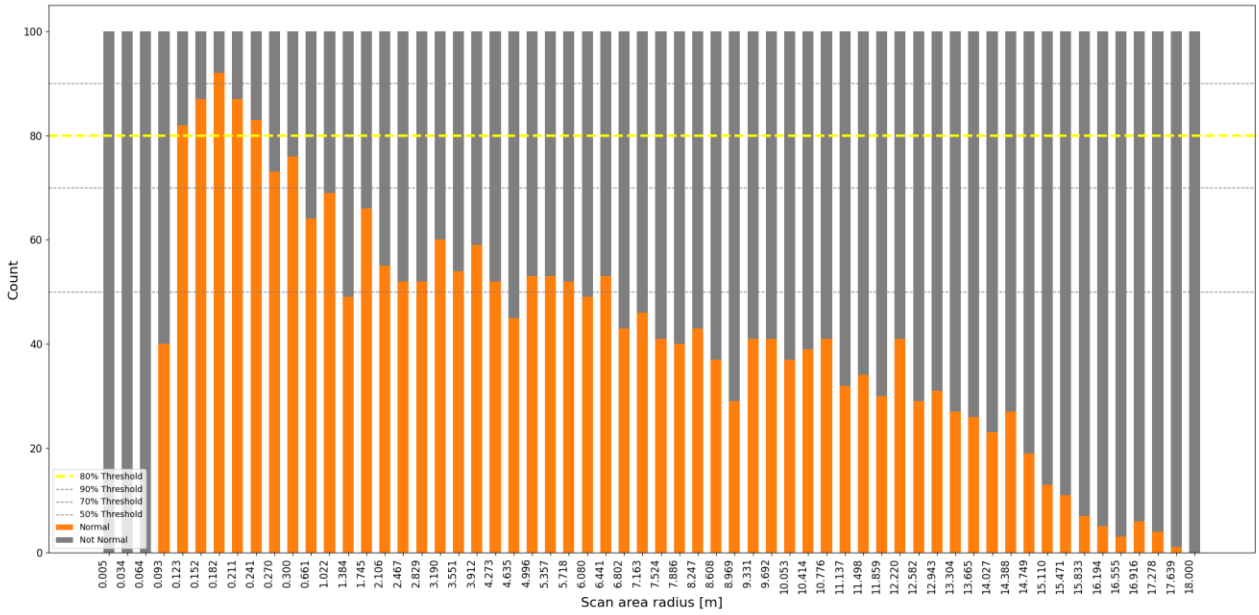


Figure 17 Shapiro-Wilk test results for Bristol channel outcrop, after the radii expansion between 0.005 m and 0.31 m.

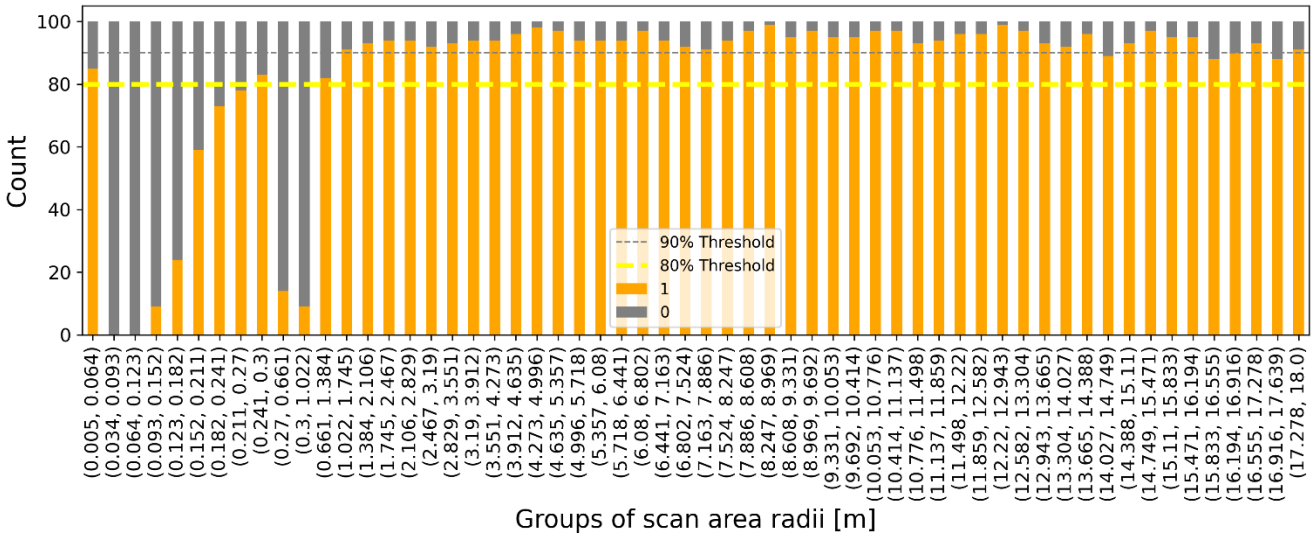


Figure 18 Levene test results for Bristol channel outcrop, after the radii expansion between 0.005 m and 0.31 m. Group size = 3.

The size-three groups 0.211–0.27 m and 0.241–0.3 m can both be considered as the REA range (Fig. 17, 18, 19, Tab. 2). This is followed by a sharp drop in acceptances and the subsequent onset of the "homogeneous" behaviour described earlier. The drop in acceptances observed around 0.3 m may be related to the change in sampling step size. Prior to 0.3 m, the sampling step is densified and remains close to 0.03 m, whereas beyond 0.3 m it increases to 0.3 m (Fig. 17, 19). This pattern is particularly evident in the residual plot shown in Figure 19C. In contrast, Figure 19D illustrates another rejection interval in which the scan area radii are equally spaced. In Figure 19E, which corresponds to the first interval excluding the 0.3 m radius, an increase in acceptances above the 80% threshold is observed. In this context, 0.3 m appears to mark the upper limit of the REA and represents the boundary between the classical "REA behaviour" discussed in this work and a regime characterized by qualitatively consistent distributions.

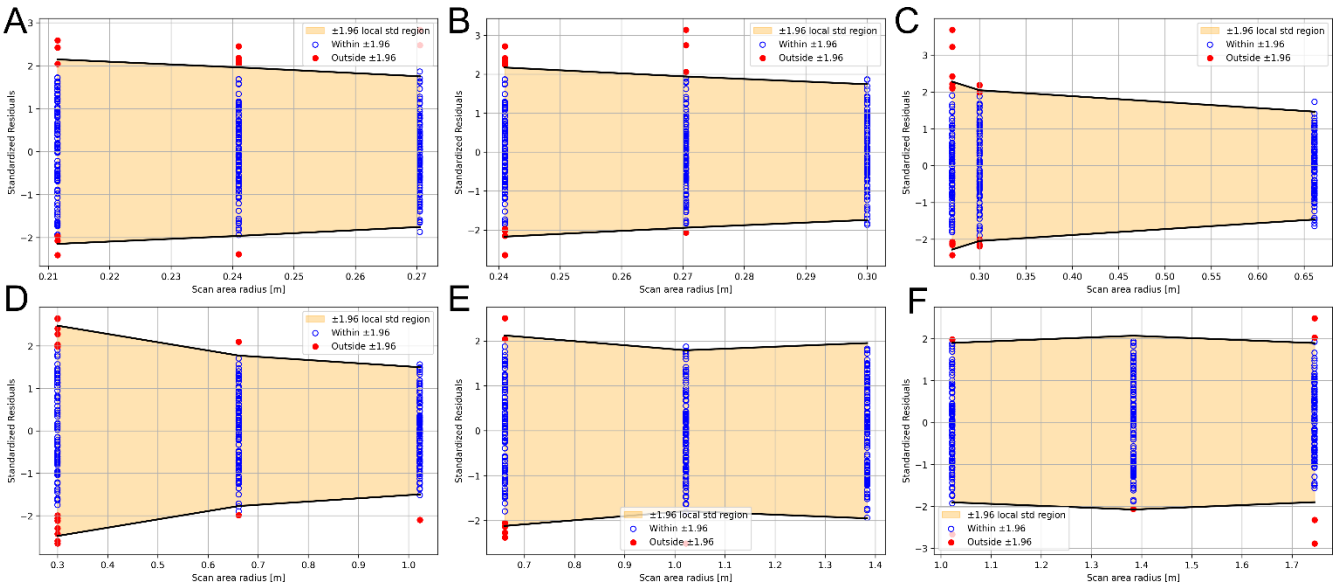


Figure 19 Residual plots for Lilstock outcrop. (A) Scan area radii from 0.211 m to 0.27 m, (B) Scan area radii from 0.241 m to 0.3 m, (C) Scan area radii from 0.27 m to 0.661 m, (D) Scan area radii from 0.3 m to 1.022 m, (E) Scan area radii from 0.661 m to 1.384 m, (F) Scan area radii from 1.022 m to 1.745 m. Figures (A) and (B) correspond to the identified REA range; Figures (C) and (D) capture the drop in acceptances; and Figures (E) and (F) represent the beginning of the "homogeneous" zone.

	ANOVA acceptances	ANOVA rejections	Test result
0.211 to 0.27 m group	95	5	accepted
0.241 to 0.3 m group	96	4	accepted

Table 2 ANOVA test results for Pontrelli quarry Outcrop.

5 Discussions

5.1 Acceptance thresholds

Both Shapiro-Wilk, Levene and ANOVA test are repeated for one hundred different realizations, with a significance level set at 0.05. Starting from the Shapiro–Wilk test, if the P_{21} values collected with scan areas of a given radius follow a pristine normal distribution, then the outcome of this "meta-analysis" should reflect the chosen significance level. In other words, around 95 out of 100 realisations should fail to reject the null hypothesis, while the remaining ~5 out of 100 are expected to incorrectly reject it due to type I error. Following this line of thought, it would be natural to think of setting the acceptance threshold at 95%.

From another perspective, the dominance of the normal model among real-world phenomena has been hugely debated (Micceri, 1989). In most cases, the so-called Data Generating Process (DGP), that is the underlying process that produce the observed data (or the errors), is unknown. Moreover, the test assesses whether a 'sample', defined

as a fixed set of realizations, appears to be drawn from a normal distribution, rather than evaluating an actual continuous random variable (Rochon et al., 2012). These considerations suggest that applying the testing procedure with the expectation of a precise outcome (e.g., a strict 95% significance threshold) may lead to misinterpretation. Although no prior information is available regarding the error distribution of P_{21} within the REA range, both exploratory data analysis and REA theory indicate that a symmetrical distribution can be expected (Fig. 4, 5, 6). Normality is therefore not an inherent property of the P_{21} values but a working hypothesis; the data should be 'normal enough' for the assumption to hold. In this study, an 80% acceptance threshold is adopted to consider P_{21} values asymptotically normal. At this threshold, the distribution closely resembles, though does not fully conform to, a normal distribution, allowing reasonable assumptions to be made about its overall shape.

This is also true for homoscedasticity and ANOVA analysis. These tests are built such that they are considered accepted if two groups have the exact same variance (Levene) or mean (ANOVA). However, in the context of natural phenomena, expecting this condition to be consistently met across realizations is unrealistic.

5.2 Sample size effect on statistical test results

The power of a statistical test, that is the probability of a correct rejection of the null hypothesis (H0), is directly related to the sample size (Kozak and Piepho, 2018). This means that at low sample size there is a greater probability of accepting a false null hypothesis (type II errors), which translates into a higher acceptances count. Increasing the sample size will progressively reduce the probability of making type II errors, to the point where, at sufficiently large sample sizes, even the smallest departure from the null hypothesis will be detected and consequently H0 will be rejected.

This arises question especially when testing the theoretical distribution of a sample. For instance, given a distribution ρ which is only slightly more or less peaked than the normal distribution, for most practical purposes, we could use the normal distribution as a model for any of its samples. However, if we draw a sufficiently large sample from ρ, and test its normality, we are bound to reject this hypothesis.

This behaviour makes the outcome of statistical tests strongly dependent on the sample size, meaning that by adjusting the sample size it becomes possible to accept or reject a null hypothesis which is only "slightly" false (and therefore acceptable for practical use). This has led to a decline in confidence in the use of formal statistical tests and a renewed emphasis on graphical approaches (e.g., probability plots), which, although more qualitative, allow the data to be assessed more directly (Kozak and Piepho, 2018). At the same time, when dealing with large datasets and multiple realizations, checking every diagnostic plot becomes time-consuming, and identifying variations in data behaviour or setting boundaries (such as defining the REA range) is not straightforward. For example, considering the Pontrelli quarry — the smallest of the case studies in terms of the number of scan area radii — there

are 26 unique radii. This results in 26 normal probability plots per realization, and therefore 2,600 plots when 100 realizations are considered. For the Levene test, 23 residual plots are produced for size-three groups, 22 for size-four groups, and so on, for a total of 171 residual plots if the analysis stops at size-eleven groups. This number must then be multiplied by the number of realizations. Checking every single plot, albeit possible, is inefficient. In this situation, we cannot exclude the possibility of using formal statistical tests, at least as screening tools, to select a range of candidate scan area radii to be cross validated with diagnostic plots. The advantage of P_{21} data is that the sample size can be arbitrarily modified to qualitatively assess the effect of sample size on statistical tests results. We performed this test on the Pontrelli case study, which, given the reduced computational time compared to the Bristol channel, allowed us to increase the sample size up to n = 500. Results of the Shapiro–Wilk test are shown in Figure 20 for sample sizes of n = 30 (Fig. 20A), n = 100 (Fig. 7), and n = 500 (Fig. 20B). For n = 30, acceptance rates are relatively high across all scan area radii, but start to decline markedly (below 50%) at radii larger than approximately 14 m. For n = 500, almost all tests are rejected, with only a few isolated acceptances around 7m, 8m, and 9m. At n = 100, the test produces a well-defined response, with distinct zones of acceptance and rejection that are consistent with both the diagnostic plots and the exploratory data analysis. In support of this choice of sample size, the effect of sample size on the Shapiro–Wilk W statistic has been investigated by Souza et al., (2023). Although their study was conducted on a different dataset, it employed the same sampling strategy used in this work. Their results indicate that the W statistic fully stabilizes for sample sizes around n = 75–100.

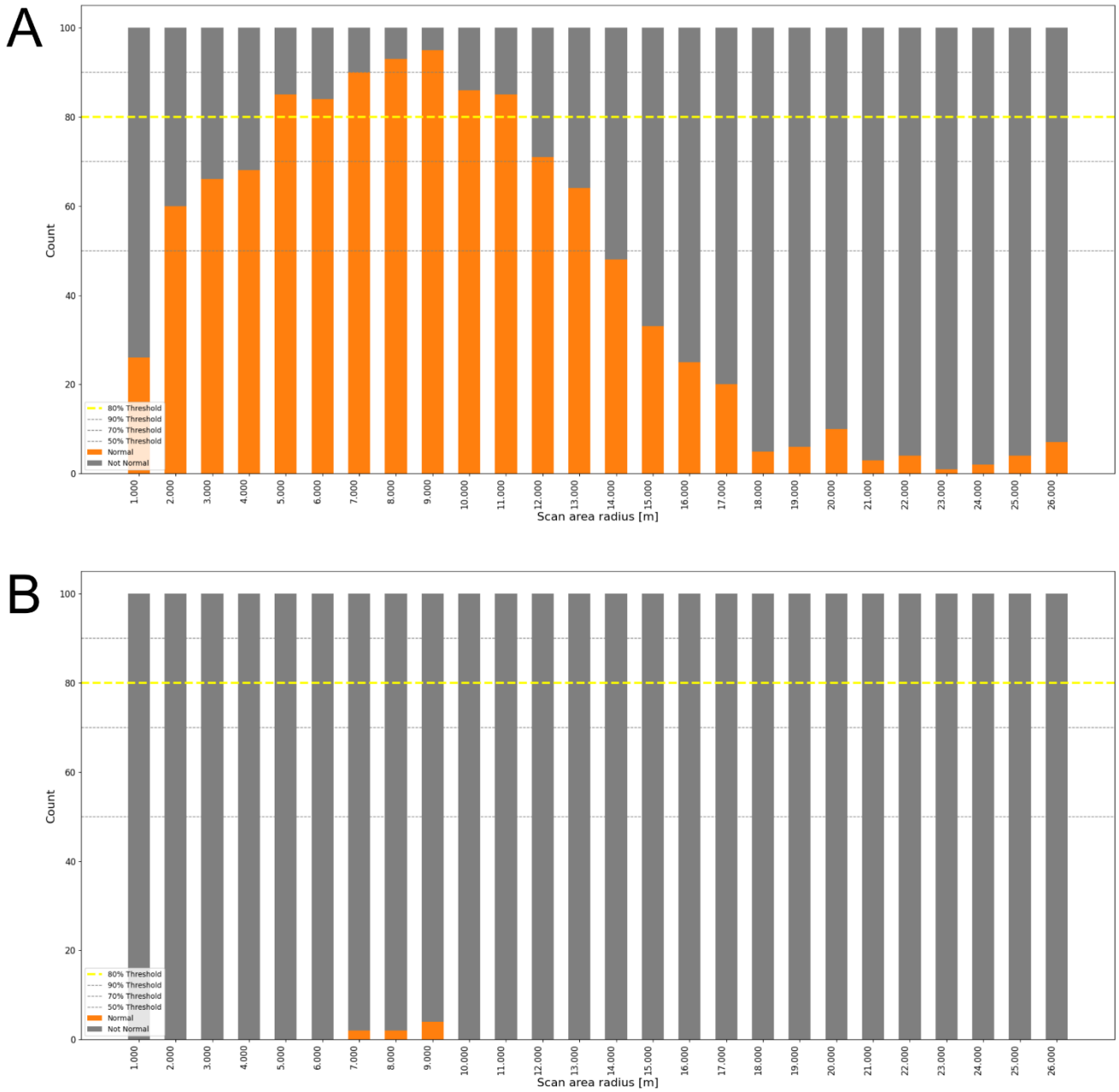


Figure 20 Shapiro-Wilk test results for Pontrelli quarry outcrop. (A) n = 30, (B) n = 500.

We do not have a decisive answer or a quantitative method for determining an ideal sample size. Moreover, even the definition of "ideal sample size" is not straightforward. One interpretation could be the sample size that maximizes the number of acceptances; however, reducing the sample size also increases the likelihood of type II errors. Another interpretation might be the sample size that maximizes test power. Yet, as we have shown, beyond a certain sample size this approach no longer yields meaningful results. Philosophically speaking, it is also

questionable to pursue perfect agreement with the assumptions of normality or homogeneity of variances, as already discussed in the previous section. In two different case studies, which have markedly different characteristics, it has been seen that the response produced by the tests at n = 100 corresponds to statistically significant peaks of acceptance and well-defined rejection zones, consistent with the diagnostic plots. Particular attention should be paid to radii with acceptance rates only slightly below the 80% threshold. The statistical power of normality tests depends on sample size, and for large samples it becomes possible to reject the null hypothesis even when the data are still reasonably consistent with a normal model. In such cases, a formal rejection does not necessarily imply that the dataset should be discarded, as the outcome may simply reflect the sensitivity of the test rather than a substantive deviation from normality. This is precisely where diagnostic plots become essential, as they provide a visual benchmark to contextualize the statistical result.

5.3 Effect of radii discretization strategy

Another variable left to the user's choice is the number and stepping strategy for scan area radii. The maximum radius depends on the outcrop size, while the minimum radius should be small enough to capture the discontinuous behavior of the fracture network in analysis. How intermediate radii are collected in between the maximum and minimum radii is a trial-and-error procedure that starts from an initial guess. Starting from the Pontrelli quarry, as an initial guess, P_{21} values were collected with scan area radii spaced at 1m intervals. A reasonable question arises as to whether decreasing the spacing between radii would lead to significant changes in the test results. By increasing the number of sampled scan area radii to 50 (Fig. 21), it appears that the overall trend of acceptances remains unchanged, and no anomalous behaviour emerges from the denser sampling. This means that keeping a sampling step of 1m or halving it does not produce significant results aside from increasing the computational time.

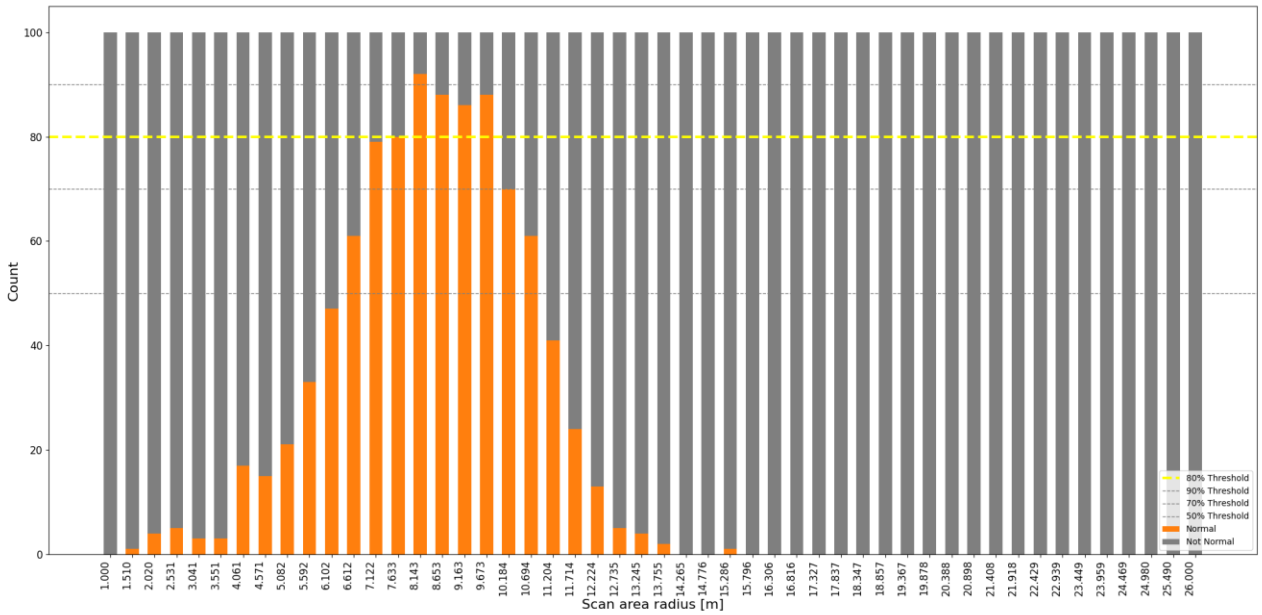


Figure 21 Shapiro-Wilk test results for Pontrelli quarry outcrop with 50 different scan area radii.

In contrast, for the Lilstock outcrop case study, the initial sampling consisted of 50 different scan area radii, as the higher fracture density suggested that P_{21} distributions might vary more substantially. Using these 50 radii, an anomalous behaviour at 0.005 m was detected, prompting a local densification of radii to investigate it further. As a cross-check, the analysis was repeated using an equal spacing of 1 m, as done for the Pontrelli quarry (Fig. 22). In this case, although the overall trend of acceptances remained unchanged, the anomaly could not be detected.

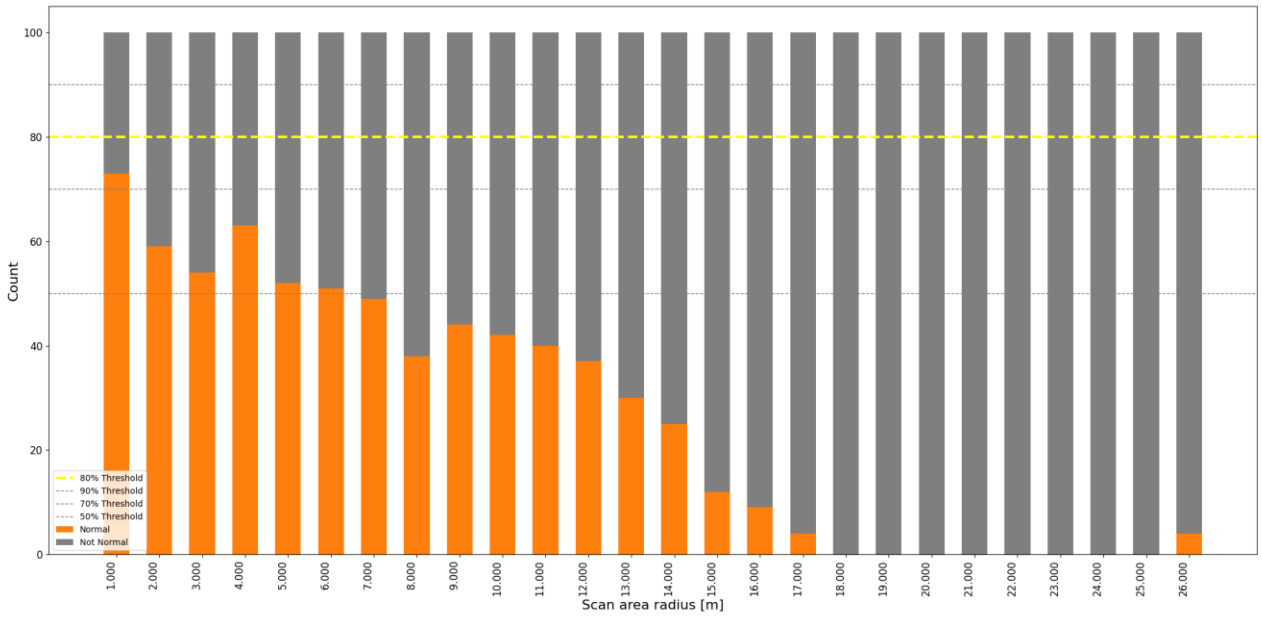


Figure 22 Shapiro-Wilk test results for Bristol channel outcrop for scan area radii with an equal spacing of 1m.

Choosing the appropriate number of scan area radii is not straightforward, as it depends on both the size of the outcrop and the characteristics of the fracture network. No universal strategy exists to define the optimal number of radii. In our case studies, we observed that starting with a relatively high number of radii is generally preferable: in the worst-case scenario, as for the Pontrelli quarry, the additional radii are simply redundant and do not affect the results, while in other situations, such as the Bristol Channel, a higher initial number allows detection of anomalies that might otherwise be missed. Alternative strategies, such as sampling radii according to a logarithmic or geometric series, could also be explored depending on the fracture network characteristics and the objectives of the analysis.

5.4 REA upper limit vs representativity limit

In the introduction section, REA has been defined as a range with a lower limit corresponding to the discontinuum-continuum transition and an upper limit that represents another transition to discontinuous behavior. In the context of a natural outcrop with limited extent, the detected upper limit is constrained by the overall outcrop size and the geometry of the interpretation boundary. For the Pontrelli quarry, when the scan area radius exceeds approximately 14 m (Fig. 3), a strong overestimation emerges relative to the normal model, producing left-tailed distributions characterized by a few small values. At this stage, the variance approaches zero and the P_{21} values converge toward a single value, the only one that can be sampled, as the scan areas become increasingly restricted by the boundary size and shape. The observed decrease in acceptance rates for the Shapiro–Wilk and Levene tests is therefore linked

to a loss of measurement representativity rather than to the detection of the true upper limit of the REA range. Nevertheless, this representativity limit provides essential information, as it defines the largest representative scan area that can be extracted from the dataset. In the Bristol Channel case study, we observed a somewhat different situation. Although REA conditions are achieved at very small scan area radii, a contrasting behaviour emerges at larger radii, where homoscedasticity acceptances increase again and probability plots display homogeneous distributions. It is difficult to determine whether the REA range holds any real significance, given that its upper and lower limit values are very close and that identifying them required a further increase in the number of sampled radii. Further case studies would be needed to investigate this behaviour more thoroughly.

5.5 Spatial autocorrelation of P21 data and applicability of ANOVA analysis

Spatial autocorrelation is a property of random variables taking values at pairs that are more similar or less similar than expected for randomly associated pairs of observations (Legendre, 1993). In the context of fracture networks and their related parameters, autocorrelation is given by, on the one hand, the underlying mechanical laws that guided the fracture generating process, and at a bigger scale by the tectonic and structural context and the presence of large-scale structures. Therefore, in a fully developed fracture network, the position of a certain fracture is not independent from the others surrounding it. This leads to the partial predictability of the P_{21} value collected with two adjacent scan areas, and the consequently violation of the independence hypothesis of the ANOVA test (Griffith, 1978, 1992). It is therefore important to contextualize the methodology presented within the problem of spatial autocorrelation and understand the conditions under which it can be applied.

The ANOVA independence assumption requires both independence between groups (e.g., P_{21} data from 1m vs. 2m scan radii) and independence of observations within each group. Regarding independence between groups, spatial autocorrelation would pose a significant concern if the study factor were the geographic position of the scan areas rather than the scan area radius. This means checking whether the mean of P_{21} variable would undergo significant changes when the position of the scan areas changes. In such experimental designs, P_{21} values collected within clusters, anticlusters, or spatial gradients do not provide entirely independent information due to the inherent spatial autocorrelation of fractures in the study area (Legendre, 1993). In the presented methodology, the factor taken into consideration is the scan area radius and every group of data is composed of scan areas sampled from the whole region considered in the analysis, every time the radius increases. This implies that the position is not taken into consideration. The same logic cannot be applied to the within-group independence assumption. A random sampling strategy with replacement does not guarantee independence if two scan areas are located within a spatially autocorrelated zone. While Spatial Independent Range Sampling (SIRS, Xie et al., 2021) techniques have been developed to address spatially autocorrelated data (cit.), our application faces a unique constraint: P_{21} is calculated over an area rather than at a discrete point. Consequently, as the scan radius increases, the likelihood of spatial

overlap between windows rises, potentially introducing redundancy or reducing the number of truly independent observations. In a situation where it is not possible to obtain both an equal sample size and an independent sample, a suggested solution is running the analysis on spatially homogeneous zones, identifying clustered or anticlustered areas and analyzing them separately, de facto removing the autocorrelation from the problem. The problem of spatial analysis of fracture arrangement has been widely addressed in scientific literature, and clusters or anticlusters can be identified through different approaches, for example the Ripley K-function (Shakiba et al., 2022), or cumulative distribution function of fracture spacing (Bistacchi et al., 2020), among others.

5.6 Relevance of a multiparametric methodology for REA definition

In this paper, we presented a multiparametric methodology based on the mean, variance and shape evaluation of the P_{21} distribution to detect the REA range, starting from data collected from outcrops large enough to distribute scan areas with radius that span some order of magnitude in radius. Compared to several studies in the scientific literature (Esmaieli et al., 2010; Zhang et al., 2012; Huang et al., 2020), we chose to calculate the REA rather than the REV in order to remain consistent with the deterministic data obtained from the TS-DOM. The determination of the REV depends not only on the geometric parameters of the fracture network but also on its spatial distribution. Using stochastic models to generate a volume for REV estimation actually replaces the natural fracture-generating process with the spatial point process assumed in the stochastic model. Most commercial and open-source tools for generating 3D stochastic DFN models rely on a Poisson point process, where the position of each fracture centroid is independent of those already generated. Consequently, calculating the REV from a stochastic DFN corresponds to evaluating the REV of a different, model-derived fracture network, rather than that of the natural outcrop.

When it comes to methods that estimate the REV from outcrop data, several studies propose defining the REA value using a multiplier associated to either the mean trace length (Li and Zhang, 2011), the average spacing (Zeeb et al., 2013), the average block size (Rohrbaugh Jr. et al., 2002) and the maximum spacing (Pariseau et al., 2008).

Applying these methods to our case studies, the multipliers proposed by Li and Zhang (2011) can only be applied to connected networks. Consequently, they cannot be applied to Pontrelli quarry, as the REA there is calculated on a single fracture set. Assuming Lilstock outcrop represents a highly connected network, the proposed multiplier is 2.8. Using the mean trace length of the fracture sets reported by Passchier et al. (2021), the average fracture length in Lilstock outcrop is approximately 7 m, resulting in a circle with a diameter of 19.6m, yielding an equivalent REA radius of about 10 m.

Rohrbaugh Jr. et al. (2002) suggested that the REA should exceed the average block size. Using the polygon shapefile provided by Prabhakaran et al. (2021), the average block size in Lilstock outcrop is 0.05 m², corresponding

to a circle with an equivalent radius of 0.13 m. In this case, the REA identified with our method aligns well with this criterion.

Zeeb et al. (2013) proposed that the REA should exceed the average fracture spacing. For Pontrelli quarry Set 1, Benedetti et al. (2025) report an average spacing of 1.55 m. Here, our method identifies a considerably larger REA, ranging from 7 m to 10 m of scan area radius.

Pariseau et al., (2008) proposed that the REV is reached at roughly twice the maximum total fracture spacing for networks with multiple sets. For Lilstock outcrop, the largest block size is 0.5 m², yielding an equivalent REV radius of 0.56 m, which approximately corresponds to the onset of homogeneous behaviour in the P_{21} values.

While these approaches have been tested across multiple case studies, they remain less informative than multi-parametric methods, which allow the data to be explored from different perspectives. For instance, in the Pontrelli case study, it is possible to observe how the distribution shape evolves, identify the upper limit, and understand the underlying causes. Moreover, such an approach can highlight anomalous behaviours, as observed in the Bristol Channel case study. In the end, the outcome may be similar, but it is achieved through a more informed and approach rather than a 'black box' one.

5.7 Comparison of REA range between case studies

We applied the methodology to two markedly different natural case studies (Fig. 2): the Pontrelli quarry outcrop, where a single fracture set is consistently detectable across the entire surface, and the Lilstock outcrop, which instead exhibits an extremely dense network composed of eight distinct fracture sets (Passchier et al., 2021). This difference is also reflected in the obtained results. In Pontrelli quarry we identified a REA range between 7 m and 10 m. The upper limit is influenced by the limited size of the outcrop, thus what is identified as an "upper limit" is instead a representativity limit, that is nonetheless and important information to keep in consideration. In the Bristol Channel case study, the detected REA range is between one and two orders of magnitude smaller than that of the Pontrelli quarry. This difference can be due to several factors. First, the Lilstock network is significantly denser — even when compared with those portions of the Pontrelli quarry where all fracture sets are visible (Casiraghi et al., 2025), Lilstock still exhibits a much higher overall intensity. Second, the employed scan areas are circular windows, an isotropic shape. Larger radii are required to reach representativity in a network dominated by a single, consistently oriented fracture set, as occurs at Pontrelli. In contrast, highly connected and multi-directional networks, such as at Lilstock, stabilize much earlier. We would likely obtain a smaller lower bound for the REA if appropriately oriented anisotropic sampling windows were used. Another difference is observed in the behaviour of the test results beyond the REA range upper limit. In the Pontrelli quarry fracture network, acceptances exhibit a progressive decline, eventually reaching zero. This applies to both the Shapiro–Wilk and Levene tests. The effect

arises because, as the scan radius increases, the placement of sampling windows becomes increasingly constrained by the interpretation boundary. Consequently, the largest scan areas tend to overlap in nearly the same locations, reducing spatial variability. The effect is further accentuated by the presence of no-data zones, which limit positioning of the scan areas even more (Fig. 3). In Lilstock, the Levene test shows a very different trend (Fig. 18): after an abrupt decrease in acceptances at small radii, the equality of the variance is consistently accepted again for larger scan areas. This occurs even though the largest windows are also constrained by the interpretation boundary. From the results of the analysis, it appears that the distributions of P_{21} at larger scan area are comparable in shape and parameters through the scan area radii, albeit not normal, in the sense in which we have defined normality in this work. It is difficult to understand if the REA range defined following the definition described in this paper is significant with respect to the overall behaviour of homoscedasticity, normality and mean P_{21} trend.

6 Conclusions

In this work, we addressed the problem of REV/REA determination for the P_{21} parameter in natural fractured outcrops. We proposed defining the REA as a range, characterized by both a lower and an upper limit. This definition is particularly relevant in the context of reservoir modelling, as it allows to define a range of cell size resolutions, from the one associated with the lowest computational cost (upper limit) to the lower limit that defines the smallest representative cell size. To calculate the REA range, we applied the ANOVA test, where the underlying assumptions are verified both with formal statistical tests and diagnostic plots.

The applicability of the method was demonstrated on two case studies with distinct characteristics. In the Pontrelli quarry, the analysis was performed on a single fracture set. The REA range was defined following the hypothesized definition. It was observed that its upper limit does not correspond to a true REA upper limit but is related to a loss of representativity due to the scan area radii size approaching the interpretation boundary size. Beyond this limit, the P_{21} distributions become left-tailed. In the Bristol Channel case study, a different behavior is observed. Even though REA conditions were achieved at small scan area radii, at larger radii the P_{21} distributions are comparable in terms of shape and distributions parameters. This is reflected in the probability plots and the resurgence of homoscedasticity acceptances. This suggests that, in some natural systems, the behavior of P_{21} may be inherently homogeneous, and therefore not adequately described by classical REA/REV definitions.

The proposed method combines multiple statistical tests with diagnostic plots, offering a multi-parameter framework for REA estimation. Rather than relying on a single numerical output, it enables the user to explore the reasons behind different statistical behaviors, enhancing interpretability and reducing the “black box” aspect often associated with statistical tests.

Some aspects of the procedure remain user-dependent, notably the sample size and the discretization strategy for the scan area radii. Both parameters directly affect the outcome of formal statistical tests. However, given that it is always possible to extract additional scan areas it is possible, at least, to qualitatively evaluate the sample size effect on the statistical tests outcome. As for the radii discretization, although there is not a quantitative rule, our results indicate that using a larger number of radii is preferable. This approach increases computational cost but does not introduce significant drawbacks, while it may reveal anomalous behaviors that would otherwise remain undetected.

General conclusions

In conclusion, the main goal of this thesis was to develop quantitative methodologies to characterize every important fracture network parameter towards more representative stochastic models.

In the first chapter we presented a comprehensive workflow for geometrical characterization of fracture networks from different types of DOMs. Orientation parameters are derived through a semi-automatic workflow. The subjectivity in defining the fracture sets is reduced by the implementation of a clustering algorithm, and the fitted distribution is tested with a formal statistical test. Topological relationships are quantified taking into account the interpretation boundary to identify B-nodes and removing them from CI calculation. Censored fractures are automatically classified. The backbone of the network is also provided. A step ahead with respect to "standard topological analysis" is made through the directional topology, where every node is associated with the corresponding fracture set (I nodes) or sets (X and Y nodes). Fracture length and height distributions are corrected for censoring bias, The best model is defined through a graphical approach and a ranking system based on statistical distances. H/L ratio, under clearly stated and testable assumptions, is estimated by regression analysis. P21 REA is defined through a qualitative multi-parameter approach that does not require to validate any specific underlying assumptions.

The second chapter demonstrates that censoring bias systematically leads to the underestimation of fracture length and shows that commonly adopted approaches in the literature (e.g., ignoring censored data, removing censored observations, or adopting circular scanline assumptions) are not suitable for stochastic modelling. Survival analysis provides an innovative framework because it explicitly incorporates the relationship between observed lengths and their censoring mechanism, does not rely on geometric assumptions, and directly estimates parameters from the measured variable of interest. The results show that even low censoring fractions (≈9%) already introduce a significant underestimation of the mean and variance in classical approaches, while survival-based estimates remain stable and closer to the natural distribution even at high censoring levels. A combined use of PIT plots and statistical distance metrics enables objective model selection. The full workflow was successfully validated across multiple geological scenarios with variable censoring intensity.

The third chapter introduces a revised definition of the Representative Elementary Area (REA) for fracture intensity (P21), conceptualized as a range bounded by both a lower and an upper limit rather than a single threshold. This definition is particularly relevant for reservoir modelling, where the upper limit represents the biggest representative cell size, associated with the lowest computational cost. The REA range is quantified through ANOVA test, with underlying assumptions verified using a combination of formal statistical tests and diagnostic plots. The proposed framework is multi-parametric, emphasizing interpretability over a black-box outcome and enabling users to diagnose the causes of statistical behavior. Some procedural aspects remain user-dependent, especially the number

of scan areas and the discretization of radii, but their influence can be qualitatively assessed, and using a higher number of radii generally improves robustness without introducing significant drawbacks.

Future prospects

What we were unable to tackle within the scope of this thesis is stochastic modelling based on the obtained parameters.

Adopting a stochastic approach to upscale fracture parameters is fundamental as fracture networks analyzed in outcrops represent a single deterministic configuration, which is unlikely to be replicated in space as it is. Calculating parameters from the fracture networks, means obtaining the "blueprint" of that specific configuration, which will be used by the stochastic model generate discrete fracture network models that mimic the analyzed fracture network. Stochastic models can be addressed from different points of view, on one hand there are statistics-based models, on the other hand there are the so called "genetic DFNs" that are based on geomechanical laws and tries to imitate the growth process of fractures (Davy et al., 2013).

When it comes to reservoir scale models, only statistics-based DFNs can be effectively applied, given that genetic DFNs are computationally demanding. Statistic-based models come with a low computational cost, and they are the models implemented in most commercial and open-source solutions. These models require fracture parameters and a stopping criterion to end the simulation. Fractures are generated in a 3D volume with a random spatial distribution, according to a Poisson point process. The geometrical properties of each fracture are drawn from parametric length and orientation distributions, and fracture height is generally controlled by a fixed height/length ratio. The simulator generates fractures until a target fracture intensity $P32$ (Dershowitz and Herda, 1992)is reached in the simulation volume. Connectivity or any other form of spatial organization cannot be taken into account in these models due to limitations of the Poisson distribution, that is specifically based on the assumption of spatial independence between fractures (Manzocchi, 2002). Modern approaches have been developed in the last years to try and solve this fundamental limitation, for instance controlling clustering of fractures by means of the Ripley's K function (Shakiba et al., 2024), or including attractive vs. repulsive spatial and directional processes controlled by statistical and/or pseudo-mechanical parametrizations (Bonneau et al., 2013; Davy et al., 2013; Bonneau et al., 2016) but a satisfactory solution has yet to be found, especially in 3D. Sometimes also "deterministic" DFNs are used, but the possibility of creating such models is limited to structures that can be imaged in 3D seismics, i.e. meso-scale faults larger than some hundred meters and characterized by an offset that results in a contrast in seismic impedance, or to small-scale sample that can be scanned with a Micro-CT apparatus.

All of this highlights the current gap between the level of detail that can be extracted from outcrop-based datasets and the ability of stochastic DFN models to reproduce those same parameters in 2D or 3D space.

Applications to real case studies

During the course of the PhD project an external collaboration with the University of Zagreb and the Croatian Geological Survey was established, to perform the characterization and consequent stochastic modelling of an outcrop analogue for a thermal aquifer, the Daruvar aquifer. In this context, the association between structural data and hydrogeological data obtained through pumping test allowed to assess the effect of the fracture network on the porosity and permeability of the aquifer. The resulting paper can be found at the following DOI: https://doi.org/10.4154/gc.2024.11

References

Ackermann, R. V. and Schlische, R. W.: Anticlustering of small normal faults around larger faults, Geology, 25, 1127–1130, https://doi.org/10.1130/0091-7613(1997)025<1127:AOSNFA>2.3.CO;2, 1997.

Agliardi, F., Crosta, G. B., Meloni, F., Valle, C., and Rivolta, C.: Structurally-controlled instability, damage and slope failure in a porphyry rock mass, Tectonophysics, 605, 34–47, https://doi.org/10.1016/j.tecto.2013.05.033, 2013.

Agliardi, F., Zanchetta, S., and Crosta, G. B.: Fabric controls on the brittle failure of folded gneiss and schist, Tectonophysics, 637, 150–162, https://doi.org/10.1016/j.tecto.2014.10.006, 2014.

Agliardi, F., Dobbs, M. R., Zanchetta, S., and Vinciguerra, S.: Folded fabric tunes rock deformation and failure mode in the upper crust, Sci. Rep., 7, 15290, https://doi.org/10.1038/s41598-017-15523-1, 2017.

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Autom. Control, 19, 716–723, https://doi.org/10.1109/TAC.1974.1100705, 1974.

Anderson, D. R. and Burnham, K. P.: Avoiding Pitfalls When Using Information-Theoretic Methods, J. Wildl. Manag., 66, 912–918, https://doi.org/10.2307/3803155, 2002.

Anderson, T. W. and Darling, D. A.: A Test of Goodness of Fit, J. Am. Stat. Assoc., 49, 765–769, https://doi.org/10.1080/01621459.1954.10501232, 1954.

Andersson, J., Shapiro, A. M., and Bear, J.: A Stochastic Model of a Fractured Rock Conditioned by Measured Information, Water Resour. Res., 20, 79–88, https://doi.org/10.1029/WR020i001p00079, 1984.

Andrews, B. J., Roberts, J. J., Shipton, Z. K., Bigi, S., Tartarello, M. C., and Johnson, G.: 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, 2019.

Baecher, G. B.: Progressively censored sampling of rock joint traces, J. Int. Assoc. Math. Geol., 12, 33–40, https://doi.org/10.1007/BF01039902, 1980.

Baecher, G. B. and Lanney, N. A.: Trace Length Biases In Joint Surveys, 19th U.S. Symposium on Rock Mechanics (USRMS), 1978.

Barton, C. C., Hsieh, P. A., Angelier, J., Bergerat, F., Bouroz, C., Dettinger, M. D., and Weeks, E. P.: Physical and Hydrologic-Flow Properties of Fractures Las Vegas, Nevada—Zion Canyon, Utah—Grand Canyon, Arizona—Yucca Mountain, Nevada July 20–24, 1989, American Geophysical Union, Washington, D. C., https://doi.org/10.1029/FT385, 1989.

Bear, J.: Dynamics of Fluids in Porous Media, Soil Sci., 120, 162, https://doi.org/10.1097/00010694-19750800000022, 1975.

Beaudoin, N. E., Lacombe, O., Hoareau, G., and Callot, J.-P.: How the geochemistry of syn-kinematic calcite cement depicts past fluid flow and assists structural interpretations: a review of concepts and applications in orogenic forelands, Geol. Mag., 159, 2157–2190, https://doi.org/10.1017/S0016756822001327, 2022.

Bellian, J. A., Kerans, C., and Jennette, D. C.: Digital Outcrop Models: Applications of Terrestrial Scanning Lidar Technology in Stratigraphic Modeling, J. Sediment. Res., 75, 166–176, https://doi.org/10.2110/jsr.2005.013, 2005.

Benedetti, G., Casiraghi, S., Bertacchi, D., and Bistacchi, A. L. P.: Unbiased statistical length analysis of linear features: Adapting survival analysis to geological applications, https://doi.org/10.5194/egusphere-2024-2818, 2025a.

Benedetti, G., Casiraghi, S., Bertacchi, D., and Bistacchi, A.: Unbiased statistical length analysis of linear features: adapting survival analysis to geological applications, Solid Earth, 16, 367–390, https://doi.org/10.5194/se-16-367-2025, 2025b.

Berio, L. R., Mittempergher, S., Storti, F., Bernasconi, S. M., Cipriani, A., Lugli, F., and Balsamo, F.: Open–closed–open palaeofluid system conditions recorded in the tectonic vein networks of the Parmelan anticline (Bornes Massif, France), J. Geol. Soc., 179, jgs2021-117, https://doi.org/10.1144/jgs2021-117, 2022.

Bisdom, K., Gauthier, B. D. M., Bertotti, G., and Hardebol, N. J.: Calibrating discrete fracture-network models with a carbonate three-dimensional outcrop fracture network: Implications for naturally fractured reservoir modeling, AAPG Bull., 98, 1351–1376, https://doi.org/10.1306/02031413060, 2014.

Bistacchi, A. and Massironi, M.: Post-nappe brittle tectonics and kinematic evolution of the north-western Alps: an integrated approach, Tectonophysics, 327, 267–292, https://doi.org/10.1016/S0040-1951(00)00206-7, 2000.

Bistacchi, A., Eva, E., Massironi, M., and Solarino, S.: Miocene to Present kinematics of the NW-Alps: evidences from remote sensing, structural analysis, seismotectonics and thermochronology, J. Geodyn., 30, 205–228, https://doi.org/10.1016/S0264-3707(99)00034-4, 2000.

Bistacchi, A., Dal Piaz, G., Massironi, M., Zattin, M., and Balestrieri, M.: The Aosta–Ranzola extensional fault system and Oligocene–Present evolution of the Austroalpine–Penninic wedge in the northwestern Alps, Int. J. Earth Sci., 90, 654–667, https://doi.org/10.1007/s005310000178, 2001.

Bistacchi, A., Griffith, W. A., Smith, S. A. F., Di Toro, G., Jones, R., and Nielsen, S.: Fault Roughness at Seismogenic Depths from LIDAR and Photogrammetric Analysis, Pure Appl. Geophys., 168, 2345–2363, https://doi.org/10.1007/s00024-011-0301-7, 2011.

Bistacchi, A., Balsamo, F., Storti, F., Mozafari, M., Swennen, R., Solum, J., Tueckmantel, C., and Taberner, C.: Photogrammetric digital outcrop reconstruction, visualization with textured surfaces, and three-dimensional structural analysis and modeling: Innovative methodologies applied to fault-related dolomitization (Vajont Limestone, Southern Alps, Italy), Geosphere, 11, 2031–2048, https://doi.org/10.1130/GES01005.1, 2015.

Bistacchi, A., Mittempergher, S., Martinelli, M., and Storti, F.: On a new robust workflow for the statistical and spatial analysis of fracture data collected with scanlines (or the importance of stationarity), Solid Earth, 11, 2535–2547, https://doi.org/10.5194/se-11-2535-2020, 2020a.

Bistacchi, A., Mittempergher, S., Martinelli, M., and Storti, F.: On a new robust workflow for the statistical and spatial analysis of fracture data collected with scanlines (or the importance of stationarity), Solid Earth, 11, 2535–2547, https://doi.org/10.5194/se-11-2535-2020, 2020b.

Bistacchi, A., Massironi, M., and Viseur, S. (Eds.): 3D Digital Geological Models: From Terrestrial Outcrops to Planetary Surfaces, 1st ed., Wiley, https://doi.org/10.1002/9781119313922, 2022a.

Bistacchi, A., Mittempergher, S., and Martinelli, M.: Digital Outcrop Model Reconstruction and Interpretation, in: 3D Digital Geological Models, edited by: Bistacchi, A., Massironi, M., and Viseur, S., Wiley, 11–32, https://doi.org/10.1002/9781119313922.ch2, 2022b.

Bonneau, F., Henrion, V., Caumon, G., Renard, P., and Sausse, J.: A methodology for pseudo-genetic stochastic modeling of discrete fracture networks, Comput. Geosci., 56, 12–22, https://doi.org/10.1016/j.cageo.2013.02.004, 2013.

Bonneau, F., Caumon, G., and Renard, P.: Impact of a stochastic sequential initiation of fractures on the spatial correlations and connectivity of discrete fracture networks, J. Geophys. Res. Solid Earth, 121, 5641–5658, https://doi.org/10.1002/2015JB012451, 2016.

Bonnet, E., Bour, O., Odling, N. E., Davy, P., Main, I., Cowie, P., and Berkowitz, B.: Scaling of fracture systems in geological media, Rev. Geophys., 39, 347–383, https://doi.org/10.1029/1999RG000074, 2001.

Boro, H., Rosero, E., and Bertotti, G.: Fracture-network analysis of the Latemar Platform (northern Italy): integrating outcrop studies to constrain the hydraulic properties of fractures in reservoir models, Pet. Geosci., 20, 79–92, https://doi.org/10.1144/petgeo2013-007, 2014.

Borradaile, G.: Statistics of Earth Science Data, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-05223-5, 2003.

Brown, M. B. and Forsythe, A. B.: Robust Tests for the Equality of Variances, J. Am. Stat. Assoc., 69, 364–367, https://doi.org/10.2307/2285659, 1974.

Brown, S. R. and Bruhn, R. L.: Fluid permeability of deformable fracture networks, J. Geophys. Res. Solid Earth, 103, 2489–2500, https://doi.org/10.1029/97JB03113, 1998.

Bruna, P.-O., Straubhaar, J., Prabhakaran, R., Bertotti, G., Bisdom, K., Mariethoz, G., and Meda, M.: A new methodology to train fracture network simulation using multiple-point statistics, Solid Earth, 10, 537–559, https://doi.org/10.5194/se-10-537-2019, 2019.

Burnham, K. P. and Anderson, D. R.: Multimodel Inference: Understanding AIC and BIC in Model Selection, Sociol. Methods Res., 33, 261–304, https://doi.org/10.1177/0049124104268644, 2004.

Cacas, M. C., Ledoux, E., de Marsily, G., Tillie, B., Barbreau, A., Durand, E., Feuga, B., and Peaudecerf, P.: Modeling fracture flow with a stochastic discrete fracture network: calibration and validation: 1. The flow model, Water Resour. Res., 26, 479–489, https://doi.org/10.1029/WR026i003p00479, 1990.

Candela, T., Renard, F., Klinger, Y., Mair, K., Schmittbuhl, J., and Brodsky, E. E.: Roughness of fault surfaces over nine decades of length scales, J. Geophys. Res. Solid Earth, 117, 2011JB009041, https://doi.org/10.1029/2011JB009041, 2012.

Casiraghi, S., Benedetti, G., Bertacchi, D., Mittempergher, S., Agliardi, F., Monopoli, B., La Valle, F., Martinelli, M., Bigoni, F., Albertini, C., and Bistacchi, A.: An integrated workflow for parametrization of fracture network geometry in digital outcrop models, Solid Earth, 16, 1351–1382, https://doi.org/10.5194/se-16-1351-2025, 2025.

Cherubini, C.: A Modeling Approach for the Study of Contamination in a Fractured Aquifer, Geotech. Geol. Eng., 26, 519–533, https://doi.org/10.1007/s10706-008-9186-3, 2008.

Clauset, A., Shalizi, C. R., and Newman, M. E. J.: Power-Law Distributions in Empirical Data | SIAM Review, SIAM Rev., 51, 661–703, https://doi.org/10.1137/070710111, 2009.

Cox, D. R.: Analysis of Survival Data, Chapman and Hall/CRC, New York, 212 pp., https://doi.org/10.1201/9781315137438, 2017.

Cruset, D., Vergés, J., Muñoz-López, D., Moragas, M., Cantarero, I., and Travé, A.: Fluid evolution from extension to compression in the Pyrenean Fold Belt and Basque-Cantabrian Basin: A review, Earth-Sci. Rev., 243, 104494, https://doi.org/10.1016/j.earscirev.2023.104494, 2023.

Davis, G. H., Reynolds, S. J., and Kluth, C. F.: Structural Geology of Rocks and Regions, John Wiley & Sons, 866 pp., 2011.

Davy, P., Le Goc, R., and Darcel, C.: A model of fracture nucleation, growth and arrest, and consequences for fracture density and scaling, J. Geophys. Res. Solid Earth, 118, 1393–1407, https://doi.org/10.1002/jgrb.50120, 2013.

Davy, P., Darcel, C., Le Goc, R., Munier, R., Selroos, J.-O., and Mas Ivars, D.: DFN, Why, How and What For, Concepts, Theories and Issues, 2nd International Discrete Fracture Network Engineering Conference, 2018.

De Toffoli, B., Massironi, M., Mazzarini, F., and Bistacchi, A.: Rheological and Mechanical Layering of the Crust Underneath Thumbprint Terrains in Arcadia Planitia, Mars, J. Geophys. Res. Planets, 126, e2021JE007007, https://doi.org/10.1029/2021JE007007, 2021.

Deluca, A. and Corral, Á.: Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions, Acta Geophys., 61, 1351–1394, https://doi.org/10.2478/s11600-013-0154-9, 2013.

Dershowitz, W. S. and Einstein, H. H.: Characterizing rock joint geometry with joint system models, Rock Mech. Rock Eng., 21, 21–51, https://doi.org/10.1007/BF01019674, 1988.

Dershowitz, W. S. and Herda, H. H.: Interpretation of fracture spacing and intensity, The 33rd U.S. Symposium on Rock Mechanics (USRMS), ARMA-92-0757, 1992.

Dewez, T. J. B., Girardeau-Montaut, D., Allanic, C., and Rohmer, J.: FACETS : A CLOUDCOMPARE PLUGIN TO EXTRACT GEOLOGICAL PLANES FROM UNSTRUCTURED 3D POINT CLOUDS, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., XLI-B5, 799–804, https://doi.org/10.5194/isprs-archives-XLI-B5-799-2016, 2016.

Eberhardt, E., Stead, D., and Coggan, J. S.: Numerical analysis of initiation and progressive failure in natural rock slopes—the 1991 Randa rockslide, Int. J. Rock Mech. Min. Sci., 41, 69–87, https://doi.org/10.1016/S1365-1609(03)00076-5, 2004.

Enders, C. K.: Maximum Likelihood Estimation, in: Encyclopedia of Statistics in Behavioral Science, John Wiley & Sons, Ltd, https://doi.org/10.1002/0470013192.bsa200, 2005.

Eppes, M. C., Rinehart, A., Aldred, J., Berberich, S., Dahlquist, M. P., Evans, S. G., Keanini, R., Laubach, S. E., Moser, F., Morovati, M., Porson, S., Rasmussen, M., and Shaanan, U.: Introducing standardized field methods for fracture-focused surface process research, Earth Surf. Dyn., 12, 35–66, https://doi.org/10.5194/esurf-12-35-2024, 2024.

Esmaieli, K., Hadjigeorgiou, J., and Grenon, M.: Estimating geometrical and mechanical REV based on synthetic rock mass models at Brunswick Mine, Int. J. Rock Mech. Min. Sci., 47, 915–926, https://doi.org/10.1016/j.ijrmms.2010.05.010, 2010.

Fisher, N. I. and Best, D. J.: Goodness-of-Fit Tests For Fisher's Distribution On The Sphere, Aust. J. Stat., 26, 142–150, https://doi.org/10.1111/j.1467-842X.1984.tb01228.x, 1984.

Fisher, R.: Dispersion on a sphere, Proc. R. Soc. Lond., 217, 295–305, 1953.

Fisher, R. A.: Statistical Methods, Experimental Design, and Scientific Inference, edited by: Bennett, J. H., Oxford University Press, https://doi.org/10.1093/oso/9780198522294.001.0001, 1990.

Fisher, R. A.: Statistical Methods for Research Workers, in: Breakthroughs in Statistics: Methodology and Distribution, edited by: Kotz, S. and Johnson, N. L., Springer, New York, NY, 66–70, https://doi.org/10.1007/978-1-4612-4380-9_6, 1992.

Follin, S., Hartley, L., Rhén, I., Jackson, P., Joyce, S., Roberts, D., and Swift, B.: 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, 2014.

Forstner, S. R. and Laubach, S. E.: Scale-dependent fracture networks, J. Struct. Geol., 165, 104748, https://doi.org/10.1016/j.jsg.2022.104748, 2022.

Forstner, S. R., Corrêa, R., Wang, Q., and Laubach, S. E.: Fracture length data for geothermal applications, Energy Geosci. Conf. Ser., 1, egc1-2024–17, https://doi.org/10.1144/egc1-2024-17, 2025.

Franzosi, F., Casiraghi, S., Colombo, R., Crippa, C., and Agliardi, F.: Quantitative Evaluation of the Fracturing State of Crystalline Rocks Using Infrared Thermography, Rock Mech. Rock Eng., 56, 6337–6355, https://doi.org/10.1007/s00603-023-03389-x, 2023a.

Franzosi, F., Crippa, C., Derron, M.-H., Jaboyedoff, M., and Agliardi, F.: Slope-Scale Remote Mapping of Rock Mass Fracturing by Modeling Cooling Trends Derived from Infrared Thermography, Remote Sens., 15, 4525, https://doi.org/10.3390/rs15184525, 2023b.

Gerstner, R., Maschler, A., Schneider-Muntau, B., Agliardi, F., Avian, M., Frießenbichler, M., and Zangerl, C.: The critical role of fracture propagation in the evolution of extensive, structurally preconditioned rockslides, Eng. Geol., 358, 108359, https://doi.org/10.1016/j.enggeo.2025.108359, 2025.

Gigli, G. and Casagli, N.: Semi-automatic extraction of rock mass structural data from high resolution LIDAR point clouds, Int. J. Rock Mech. Min. Sci., 48, 187–198, https://doi.org/10.1016/j.ijrmms.2010.11.009, 2011.

Gilitschenski, I., Kurz, G., Julier, S. J., and Hanebeck, U. D.: Unscented Orientation Estimation Based on the Bingham Distribution, IEEE Trans. Autom. Control, 61, 172–177, https://doi.org/10.1109/TAC.2015.2423831, 2016.

Giuffrida, A., Agosta, F., Rustichelli, A., Panza, E., La Bruna, V., Eriksson, M., Torrieri, S., and Giorgioni, M.: Fracture stratigraphy and DFN modelling of tight carbonates, the case study of the Lower Cretaceous carbonates exposed at the Monte Alpi (Basilicata, Italy), Mar. Pet. Geol., 112, 104045, https://doi.org/10.1016/j.marpetgeo.2019.104045, 2020.

Griffith, D. A.: A Spatially Adjusted ANOVA Model, Geogr. Anal., 10, 296–301, https://doi.org/10.1111/j.1538-4632.1978.tb00661.x, 1978.

Griffith, D. A.: A spatially adjusted N-way ANOVA model, Reg. Sci. Urban Econ., 22, 347–369, https://doi.org/10.1016/0166-0462(92)90034-X, 1992.

Gueguen, Y., David, C., and Gavrilenko, P.: Percolation networks and fluid transport in the crust, Geophys. Res. Lett., 18, 931–934, https://doi.org/10.1029/91GL00951, 1991.

Hadgu, T., Karra, S., Kalinina, E., Makedonska, N., Hyman, J. D., Klise, K., Viswanathan, H. S., and Wang, Y.: A comparative study of discrete fracture network and equivalent continuum models for simulating flow and transport in the far field of a hypothetical nuclear waste repository in crystalline host rock, J. Hydrol., 553, 59–70, https://doi.org/10.1016/j.jhydrol.2017.07.046, 2017.

Hancock, P. L.: Brittle microtectonics: principles and practice, J. Struct. Geol., 7, 437–457, https://doi.org/10.1016/0191-8141(85)90048-3, 1985.

Haridy, M. G., Sedighi, F., Ghahri, P., Ussenova, K., and Zhiyenkulov, M.: Comprehensive Study of the Oda Corrected Permeability Upscaling Method, in: Day 2 Wed, October 30, 2019, SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, Bali, Indonesia, D022S006R006, https://doi.org/10.2118/196399-MS, 2020.

Healy, D., Rizzo, R. E., Cornwell, D. G., Farrell, N. J. C., Watkins, H., Timms, N. E., Gomez-Rivas, E., and Smith, M.: 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, 2017.

Hoek, E.: Strength of jointed rock masses, Géotechnique, 33, 187–223, https://doi.org/10.1680/geot.1983.33.3.187, 1983.

Hoek, E. and Brown, E. T.: Practical estimates of rock mass strength, Int. J. Rock Mech. Min. Sci., 34, 1165–1186, https://doi.org/10.1016/S1365-1609(97)80069-X, 1997.

Huang, H., Shen, J., Chen, Q., and Karakus, M.: Estimation of REV for fractured rock masses based on Geological Strength Index, Int. J. Rock Mech. Min. Sci., 126, 104179, https://doi.org/10.1016/j.ijrmms.2019.104179, 2020.

Hudson, J. A. and Harrison, J. P.: Chapter 4. In Situ Stress, in: Engineering Rock Mechanics, Pergamon, Amsterdam [etc.], 1997.

Hyman, J. D., Karra, S., Makedonska, N., Gable, C. W., Painter, S. L., and Viswanathan, H. S.: dfnWorks: A discrete fracture network framework for modeling subsurface flow and transport, Comput. Geosci., 84, 10–19, https://doi.org/10.1016/j.cageo.2015.08.001, 2015.

Kalbfleisch, J. D. and Prentice, R. L.: The Statistical Analysis of Failure Time Data, John Wiley & Sons, 462 pp., 2002.

Kaplan, E. L. and Meier, P.: Nonparametric Estimation from Incomplete Observations, J. Am. Stat. Assoc., 53, 457–481, https://doi.org/10.1080/01621459.1958.10501452, 1958.

Karim, Md. R. and Islam, M. A.: Reliability and Survival Analysis, Springer Singapore, Singapore, https://doi.org/10.1007/978-981-13-9776-9, 2019.

Kaufman, L. and Rousseeuw, P. J.: Clustering by means of medoids, Stat. Data Anal. Based L1 Norm Relat. Methods, 405–416, 1987.

Kaufman, L. and Rousseeuw, P. J.: Finding groups in data: an introduction to cluster analysis, Wiley, Hoboken, N.J, 342 pp., 2005.

Kent, J. T.: The Fisher-Bingham Distribution on the Sphere, J. R. Stat. Soc. Ser. B Methodol., 44, 71–80, https://doi.org/10.1111/j.2517-6161.1982.tb01189.x, 1982.

Kim, N.: Tests based on EDF statistics for randomly censored normal distributions when parameters are unknown, Commun. Stat. Appl. Methods, 26, 431–443, https://doi.org/10.29220/CSAM.2019.26.5.431, 2019.

Kleinbaum, D. G. and Klein, M.: Survival Analysis: A Self-Learning Text, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4419-6646-9, 2012.

Kozak, M. and Piepho, H.-P.: What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions, J. Agron. Crop Sci., 204, 86–98, https://doi.org/10.1111/jac.12220, 2018.

Koziol, J. A. and Green, S. B.: A Cramér-von Mises statistic for randomly censored data, Biometrika, 63, 465–474, https://doi.org/10.1093/biomet/63.3.465, 1976.

Kulatilake, P. H. S. W. and Panda, B. B.: Effect of Block Size and Joint Geometry on Jointed Rock Hydraulics and REV, J. Eng. Mech., 126, 850–858, https://doi.org/10.1061/(ASCE)0733-9399(2000)126:8(850), 2000.

Kulatilake, P. H. S. W. and Wu, T. H.: Estimation of mean trace length of discontinuities, Rock Mech. Rock Eng., 17, 215–232, https://doi.org/10.1007/BF01032335, 1984.

Kurz, G., Gilitschenski, I., Julier, S., and Hanebeck, U. D.: Recursive Bingham Filter for Directional Estimation Involving 180 Degree Symmetry, J. Adv. Inf. FUSION, VOL. 9, 2014.

Laubach, S. E., Lander, R. H., Criscenti, L. J., Anovitz, L. M., Urai, J. L., Pollyea, R. M., Hooker, J. N., Narr, W., Evans, M. A., Kerisit, S. N., Olson, J. E., Dewers, T., Fisher, D., Bodnar, R., Evans, B., Dove, P., Bonnell, L. M., Marder, M. P., and Pyrak-Nolte, L.: The Role of Chemistry in Fracture Pattern Development and Opportunities to Advance Interpretations of Geological Materials, Rev. Geophys., 57, 1065–1111, https://doi.org/10.1029/2019RG000671, 2019.

Lawless, J. F.: Statistical Models and Methods for Lifetime Data, John Wiley & Sons, 662 pp., 2003.

Legendre, P.: Spatial Autocorrelation: Trouble or New Paradigm?, Ecology, 74, 1659–1673, https://doi.org/10.2307/1939924, 1993.

Leung, K.-M., Elashoff, R. M., and Afifi, A. A.: Censoring Issues in Survival Analysis, Annu. Rev. Public Health, 18, 83–104, https://doi.org/10.1146/annurev.publhealth.18.1.83, 1997.

Lewis, T. and Fisher, N. I.: Graphical methods for investigating the fit of a Fisher distribution to spherical data, Geophys. J. Int., 69, 1–13, https://doi.org/10.1111/j.1365-246X.1982.tb04931.x, 1982.

Li, J. H. and Zhang, L. M.: Connectivity of a network of random discontinuities, Comput. Geotech., 38, 217–226, https://doi.org/10.1016/j.compgeo.2010.11.010, 2011.

Mandl, G.: Rock Joints, Springer-Verlag, Berlin/Heidelberg, https://doi.org/10.1007/b137623, 2005.

Manzocchi, T.: The connectivity of two-dimensional networks of spatially correlated fractures, Water Resour. Res., 38, 1-1-1–20, https://doi.org/10.1029/2000WR000180, 2002.

March, R., Elder, H., Doster, F., and Geiger, S.: Accurate Dual-Porosity Modeling of CO2 Storage in Fractured Reservoirs, in: Day 3 Wed, February 22, 2017, SPE Reservoir Simulation Conference, Montgomery, Texas, USA, D031S012R004, https://doi.org/10.2118/182646-MS, 2017.

Mardia, K. V. and Jupp, P. E.: Directional statistics, J. Wiley, Chichester ; New York, 429 pp., 2000.

Marinos, V., Marinos, P., and Hoek, E.: The geological strength index: applications and limitations, Bull. Eng. Geol. Environ., 64, 55–65, https://doi.org/10.1007/s10064-004-0270-5, 2005.

Marrett, R., Gale, J. F. W., Gómez, L. A., and Laubach, S. E.: Correlation analysis of fracture arrangement in space, J. Struct. Geol., 108, 16–33, https://doi.org/10.1016/j.jsg.2017.06.012, 2018.

Martinelli, M., Bistacchi, A., Mittempergher, S., Bonneau, F., Balsamo, F., Caumon, G., and Meda, M.: Damage zone characterization combining scan-line and scan-area analysis on a km-scale Digital Outcrop Model: The Qala Fault (Gozo), J. Struct. Geol., 140, 104144, https://doi.org/10.1016/j.jsg.2020.104144, 2020a.

Martinelli, M., Bistacchi, A., Mittempergher, S., Bonneau, F., Balsamo, F., Caumon, G., and Meda, M.: Damage zone characterization combining scan-line and scan-area analysis on a km-scale Digital Outcrop Model: The Qala Fault (Gozo), J. Struct. Geol., 140, 104144, https://doi.org/10.1016/j.jsg.2020.104144, 2020b.

Mauldon, M.: 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, 1998.

Mauldon, M., Dunne, W. M., and Rohrbaugh, M. B.: 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, 2001.

Medici, G., Munn, J. D., and Parker, B. L.: Delineating aquitard characteristics within a Silurian dolostone aquifer using high-density hydraulic head and fracture datasets, Hydrogeol. J., 32, 1663–1691, https://doi.org/10.1007/s10040-024-02824-9, 2024.

Menegoni, N., Giordan, D., Perotti, C., and Tannant, D. D.: Detection and geometric characterization of rock mass discontinuities using a 3D high-resolution digital outcrop model generated from RPAS imagery – Ormea rock slope, Italy, Eng. Geol., 252, 145–163, https://doi.org/10.1016/j.enggeo.2019.02.028, 2019.

Micceri, T.: The Unicorn, The Normal Curve, and Other Improbable Creatures, Psychol. Bull., 105, 156–166, 1989.

Mittempergher, S. and Bistacchi, A.: Image Analysis Algorithms for Semiautomatic Lineament Detection in Geological Outcrops, in: 3D Digital Geological Models, John Wiley & Sons, Ltd, 93–107, https://doi.org/10.1002/9781119313922.ch6, 2022.

Moder, K.: Alternatives to F-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study), 2010.

Mooi, E., Sarstedt, M., and Mooi-Reci, I.: Market Research, Springer Singapore, Singapore, https://doi.org/10.1007/978-981-10-5218-7, 2018.

Ni, P., Wang, S., Wang, C., and Zhang, S.: Estimation of REV Size for Fractured Rock Mass Based on Damage Coefficient, Rock Mech. Rock Eng., 50, 555–570, https://doi.org/10.1007/s00603-016-1122-x, 2017.

Nicosia, U., Marino, M., Mariotti, N., Muraro, C., Panigutti, S., Petti, F. M., and Sacchi, E.: The Late Cretaceus dinosaur tracksite near Altamura (Bari, Southern Italy), Geol. Romana, 35, 237–247, 1999.

Nyberg, B., Nixon, C. W., and Sanderson, D. J.: 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, 2018.

Oda, M.: A method for evaluating the effect of crack geometry on the mechanical behavior of cracked rock masses, Mech. Mater., 2, 163–171, https://doi.org/10.1016/0167-6636(83)90035-2, 1983.

Odling, N. E.: 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, 1997.

Odling, N. E., Gillespie, P., Bourgine, B., Castaing, C., Chiles, J.-P., Christensen, N. P., Fillion, E., Genter, A., Olsen, C., Thrane, L., Trice, R., Aarseth, E., Walsh, J. J., and Watterson, J.: Variations in fracture system geometry and their implications for fluid flow in fractured hydrocarbon reservoirs, Pet. Geosci., https://doi.org/10.1144/petgeo.5.4.373, 1999.

Ortega, O. and Marrett, R.: Prediction of macrofracture properties using microfracture information, Mesaverde Group sandstones, San Juan basin, New Mexico, J. Struct. Geol., 22, 571–588, https://doi.org/10.1016/S0191-8141(99)00186-8, 2000.

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

Pahl, P. J.: Estimating the Mean Length of Discontinuity Traces, Int J Rock Mech Min Sci Geomech, 18, 221 to 228, 1981.

Panara, Y., Menegoni, N., Finkbeiner, T., Zühlke, R., and Vahrenkamp, V.: High-resolution analysis of 3D fracture networks from Digital Outcrop Models, correlation to plate-tectonic events and calibration of subsurface models (Jurassic, Arabian Plate), Mar. Pet. Geol., 167, 106998, https://doi.org/10.1016/j.marpetgeo.2024.106998, 2024.

Panza, E., Berto, C., Luzi, E., Agosta, F., Zambrano, M., Tondi, E., Prosser, G., Giorgioni, M., and Janiseck, J. M.: Structural architecture and Discrete Fracture Network modelling of layered fractured carbonates (Altamura Fm., Italy), Ital. J. Geosci., 134, 409–422, https://doi.org/10.3301/IJG.2014.28, 2015.

Panza, E., Agosta, F., Rustichelli, A., Zambrano, M., Tondi, E., Prosser, G., Giorgioni, M., and Janiseck, J. M.: Fracture stratigraphy and fluid flow properties of shallow-water, tight carbonates: The case study of the Murge Plateau (southern Italy), Mar. Pet. Geol., 73, 350–370, https://doi.org/10.1016/j.marpetgeo.2016.03.022, 2016.

Panza, E., Agosta, F., Rustichelli, A., Vinciguerra, S. C., Ougier-Simonin, A., Dobbs, M., and Prosser, G.: Meso-to-microscale fracture porosity in tight limestones, results of an integrated field and laboratory study, Mar. Pet. Geol., 103, 581–595, https://doi.org/10.1016/j.marpetgeo.2019.01.043, 2019.

Pariseau, W. G., Puri, S., and Schmelter, S. C.: A new model for effects of impersistent joint sets on rock slope stability, Int. J. Rock Mech. Min. Sci., 45, 122–131, https://doi.org/10.1016/j.ijrmms.2007.05.001, 2008.

Passchier, M., Passchier, C. W., Weismüller, C., and Urai, J. L.: The joint sets on the Lilstock Benches, UK. Observations based on mapping a full resolution UAV-based image, J. Struct. Geol., 147, 104332, https://doi.org/10.1016/j.jsg.2021.104332, 2021.

Philip, Z. G., Jennings, J. W., Olson, J. E., Laubach, S. E., and Holder, J.: Modeling Coupled Fracture-Matrix Fluid Flow in Geomechanically Simulated Fracture Networks, SPE Reserv. Eval. Eng., 8, 300–309, https://doi.org/10.2118/77340-PA, 2005.

Piaz, G. V. D., Bistacchi, A., and Massironi, M.: Geological outline of the Alps, Episodes J. Int. Geosci., 26, 175–180, https://doi.org/10.18814/epiiugs/2003/v26i3/004, 2003.

Prabhakaran, R., Urai, J. L., Bertotti, G., Weismüller, C., and Smeulders, D. M. J.: Large-scale natural fracture network patterns: Insights from automated mapping in the Lilstock (Bristol Channel) limestone outcrops, J. Struct. Geol., 150, 104405, https://doi.org/10.1016/j.jsg.2021.104405, 2021.

Rajaraman, S., Jaeger, S., and Antani, S. K.: Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images, PeerJ, 7, e6977, https://doi.org/10.7717/peerj.6977, 2019.

Renshaw, C. E.: Influence of subcritical fracture growth on the connectivity of fracture networks, Water Resour. Res., 32, 1519–1530, https://doi.org/10.1029/96WR00711, 1996.

Riquelme, A. J., Abellán, A., Tomás, R., and Jaboyedoff, M.: A new approach for semi-automatic rock mass joints recognition from 3D point clouds, Comput. Geosci., 68, 38–52, https://doi.org/10.1016/j.cageo.2014.03.014, 2014.

Riquelme, A. J., Abellán, A., and Tomás, R.: Discontinuity spacing analysis in rock masses using 3D point clouds, Eng. Geol., 195, 185–195, https://doi.org/10.1016/j.enggeo.2015.06.009, 2015.

Rizzo, R. E., Healy, D., and De Siena, L.: Benefits of maximum likelihood estimators for fracture attribute analysis: Implications for permeability and up-scaling, J. Struct. Geol., 95, 17–31, https://doi.org/10.1016/j.jsg.2016.12.005, 2017.

Rochon, J., Gondan, M., and Kieser, M.: To test or not to test: Preliminary assessment of normality when comparing two independent samples, BMC Med. Res. Methodol., 12, 81, https://doi.org/10.1186/1471-2288-12-81, 2012.

Rohrbaugh Jr., M. B., Dunne, W. M., and Mauldon, M.: Estimating fracture trace intensity, density, and mean length using circular scan lines and windows, AAPG Bull., 86, https://doi.org/10.1306/61EEDE0E-173E-11D7-8645000102C1865D, 2002.

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

Sanderson, D. J., Peacock, D. C. P., Nixon, C. W., and Rotevatn, A.: Graph theory and the analysis of fracture networks, J. Struct. Geol., 125, 155–165, https://doi.org/10.1016/j.jsg.2018.04.011, 2019.

Sari, M.: Determination of representative elementary volume (REV) for jointed rock masses exhibiting scale-dependent behavior: a numerical investigation, Int. J. Geo-Eng., 12, 34, https://doi.org/10.1186/s40703-021-00164-1, 2021.

Schultz, R. A.: Geologic Fracture Mechanics, 2019.

Schultz, R. A. and Fossen, H.: Displacement–length scaling in three dimensions: the importance of aspect ratio and application to deformation bands, J. Struct. Geol., 24, 1389–1411, https://doi.org/10.1016/S0191-8141(01)00146-8, 2002.

Sethian, J. A.: Fast Marching Methods, SIAM Rev., 41, 199–235, https://doi.org/10.1137/S0036144598347059, 1999.

Shakiba, M., Lake, L. W., Gale, J. F. W., and Pyrcz, M. J.: Multiscale spatial analysis of fracture arrangement and pattern reconstruction using Ripley's *K*-function, J. Struct. Geol., 155, 104531, https://doi.org/10.1016/j.jsg.2022.104531, 2022.

Shakiba, M., Lake, L. W., Gale, J. F. W., Laubach, S. E., and Pyrcz, M. J.: Stochastic reconstruction of fracture network pattern using spatial point processes, Geoenergy Sci. Eng., 236, 212741, https://doi.org/10.1016/j.geoen.2024.212741, 2024.

Shapiro, S. S. and Wilk, M. B.: An Analysis of Variance Test for Normality (Complete Samples), Biometrika, 52, 591–611, https://doi.org/10.2307/2333709, 1965.

Smeraglia, L., Mercuri, M., Tavani, S., Pignalosa, A., Kettermann, M., Billi, A., and Carminati, E.: 3D Discrete Fracture Network (DFN) models of damage zone fluid corridors within a reservoir-scale normal fault in carbonates: Multiscale approach using field data and UAV imagery, Mar. Pet. Geol., 126, 104902, https://doi.org/10.1016/j.marpetgeo.2021.104902, 2021.

Smirnov, N. V.: On the Estimation of the Discrepancy between Empirical Curves of Distributions for Two Independent Samples, Mosc. Univ. Math. Bull., 3–26, 1939.

Smith, S. A. F., Bistacchi, A., Mitchell, T. M., Mittempergher, S., and Di Toro, G.: The structure of an exhumed intraplate seismogenic fault in crystalline basement, Tectonophysics, 599, 29–44, https://doi.org/10.1016/j.tecto.2013.03.031, 2013.

Song, B., Zheng, N., Li, D., Chen, R., and Li, L.: Reconstructing DEM using TLS point cloud data and NURBS surface, Trans. Nonferrous Met. Soc. China, 25, 3165–3172, https://doi.org/10.1016/S1003-6326(15)63947-4, 2015.

Souza, R. R. D., Toebe, M., Mello, A. C., and Bittencourt, K. C.: Sample size and Shapiro-Wilk test: An analysis for soybean grain yield, Eur. J. Agron., 142, 126666, https://doi.org/10.1016/j.eja.2022.126666, 2023.

Stahle, L. and Wold, S.: Analysis of variance (ANOVA), Chemom. Intell. Lab. Syst., 6, 259–272, https://doi.org/10.1016/0169-7439(89)80095-4, 1989.

Staub, I., Fredriksson, A., and Outters, N.: Strategy for a Rock Mechanics Site Descriptive Model. Development and testing of the theoretical approach, 2002.

Stephens, M. A.: EDF Statistics for Goodness of Fit and Some Comparisons, J. Am. Stat. Assoc., 69, 730–737, https://doi.org/10.1080/01621459.1974.10480196, 1974.

Storti, F., Bistacchi, A., Borsani, A., Balsamo, F., Fetter, M., and Ogata, K.: Spatial and spacing distribution of joints at (over-) saturation in the turbidite sandstones of the Marnoso-Arenacea Fm. (Northern Apennines, Italy), J. Struct. Geol., 156, 104551, https://doi.org/10.1016/j.jsg.2022.104551, 2022a.

Storti, F., Bistacchi, A., Borsani, A., Balsamo, F., Fetter, M., and Ogata, K.: Spatial and spacing distribution of joints at (over-) saturation in the turbidite sandstones of the Marnoso-Arenacea Fm. (Northern Apennines, Italy), J. Struct. Geol., 156, 104551, https://doi.org/10.1016/j.jsg.2022.104551, 2022b.

Sturzenegger, M. and Stead, D.: Close-range terrestrial digital photogrammetry and terrestrial laser scanning for discontinuity characterization on rock cuts, Eng. Geol., 106, 163–182, https://doi.org/10.1016/j.enggeo.2009.03.004, 2009.

Sturzenegger, M., Stead, D., and Elmo, D.: Terrestrial remote sensing-based estimation of mean trace length, trace intensity and block size/shape, Eng. Geol., 119, 96–111, https://doi.org/10.1016/j.enggeo.2011.02.005, 2011.

Suzuki, K., Oda, M., Yamazaki, M., and Kuwahara, T.: Permeability changes in granite with crack growth during immersion in hot water, Int. J. Rock Mech. Min. Sci., 35, 907–921, https://doi.org/10.1016/S0148-9062(98)00016-3, 1998.

Taty Moukati, F., Stoica, R. S., Bonneau, F., Wu, X., and Caumon, G.: From Fault Likelihood to Fault Networks: Stochastic Seismic Interpretation Through a Marked Point Process with Interactions, Math. Geosci., 57, 115–151, https://doi.org/10.1007/s11004-024-10150-9, 2025.

Tavakkoli, M., Mohammadsadeghi, M., Shahrabadi, A., Khajoee, S., Malakooti, R., and Beidokhti, M. S.: Deterministic versus Stochastic Discrete Fracture Network (DFN) Modeling, Application in a Heterogeneous Naturally Fractured Reservoir, Kuwait International Petroleum Conference and Exhibition, https://doi.org/10.2118/127086-MS, 2009.

Tavani, S., Granado, P., Corradetti, A., Girundo, M., Iannace, A., Arbués, P., Muñoz, J. A., and Mazzoli, S.: Building a virtual outcrop, extracting geological information from it, and sharing the results in Google Earth via OpenPlot and Photoscan: An example from the Khaviz Anticline (Iran), Comput. Geosci., 63, 44–53, https://doi.org/10.1016/j.cageo.2013.10.013, 2014.

Tavani, S., Corradetti, A., and Billi, A.: High precision analysis of an embryonic extensional fault-related fold using 3D orthorectified virtual outcrops: The viewpoint importance in structural geology, J. Struct. Geol., 86, 200–210, https://doi.org/10.1016/j.jsg.2016.03.009, 2016.

Terzaghi, R. D.: Sources of Error in Joint Surveys, Géotechnique, 15, 287–304, https://doi.org/10.1680/geot.1965.15.3.287, 1965a.

Terzaghi, R. D.: Sources of Error in Joint Surveys, Géotechnique, 15, 287–304, https://doi.org/10.1680/geot.1965.15.3.287, 1965b.

Thiele, S., Grose, L., and Micklethwaite, S.: Compass: A CloudCompare workflow for digital mapping and structural analysis, 5548, 2018.

Thiele, S. T., Grose, L., Samsu, A., Micklethwaite, S., Vollgger, S. A., and Cruden, A. R.: 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, 2017.

Tukey, J. W.: Exploratory Data Analysis, Addison-Wesley Publ. Co. Read., https://doi.org/10.1002/bimj.4710230408, 1977.

Wallace, R. L., Cai, Z., Zhang, H., Zhang, K., and Guo, C.: Utility-scale subsurface hydrogen storage: UK perspectives and technology, Int. J. Hydrog. Energy, 46, 25137–25159, https://doi.org/10.1016/j.ijhydene.2021.05.034, 2021.

Wang, M., Kulatilake, P. H. S. W., Um, J., and Narvaiz, J.: Estimation of REV size and three-dimensional hydraulic conductivity tensor for a fractured rock mass through a single well packer test and discrete fracture fluid flow modeling, Int. J. Rock Mech. Min. Sci., 39, 887–904, https://doi.org/10.1016/S1365-1609(02)00067-9, 2002.

Wang, Y., Vuik, C., and Hajibeygi, H.: CO2 Storage in deep saline aquifers: impacts of fractures on hydrodynamic trapping, Int. J. Greenh. Gas Control, 113, 103552, https://doi.org/10.1016/j.ijggc.2021.103552, 2022.

Warburton, P. M.: A stereological interpretation of joint trace data, Int. J. Rock Mech. Min. Sci. Geomech. Abstr., 17, 181–190, https://doi.org/10.1016/0148-9062(80)91084-0, 1980.

Warren, J. E. and Root, P. J.: The Behavior of Naturally Fractured Reservoirs, Soc. Pet. Eng. J., 3, 245–255, https://doi.org/10.2118/426-PA, 1963.

Xia, L., Zheng, Y., and Yu, Q.: Estimation of the REV size for blockiness of fractured rock masses, Comput. Geotech., 76, 83–92, https://doi.org/10.1016/j.compgeo.2016.02.016, 2016.

Xie, D., Phillips, J. M., Matheny, M., and Li, F.: Spatial Independent Range Sampling, in: Proceedings of the 2021 International Conference on Management of Data, SIGMOD/PODS '21: International Conference on Management of Data, Virtual Event China, 2023–2035, https://doi.org/10.1145/3448016.3452806, 2021.

Yamaji A.: Genetic algorithm for fitting a mixed Bingham distribution to 3D orientations: a tool for the statistical and paleostress analyses of fracture orientations, Isl. Arc, 25, 72–83, https://doi.org/10.1111/iar.12135, 2016.

Numerical modelling of brittle fracture and step-path failure: from laboratory to rock slope scale: https://summit.sfu.ca/item/9076, last access: 27 October 2025.

Zambrano, M., Tondi, E., Korneva, I., Panza, E., Agosta, F., Janiseck, J. M., and Giorgioni, M.: Fracture properties analysis and discrete fracture network modelling of faulted tight limestones, Murge Plateau, Italy, Ital. J. Geosci., 135, 55–67, https://doi.org/10.3301/IJG.2014.42, 2016.

Zeeb, C., Gomez-Rivas, E., Bons, P. D., and Blum, P.: Evaluation of sampling methods for fracture network characterization using outcrops, AAPG Bull., 97, 1545–1566, https://doi.org/10.1306/02131312042, 2013.

Zhang, L. and Einstein, H. H.: Estimating the Mean Trace Length of Rock Discontinuities, Rock Mech. Rock Eng., 31, 217–235, https://doi.org/10.1007/s006030050022, 1998.

Zhang, L., Einstein, H. H., and Dershowitz, W. S.: Stereological relationship between trace length and size distribution of elliptical discontinuities, Géotechnique, 52, 419–433, https://doi.org/10.1680/geot.2002.52.6.419, 2002.

Zhang, W., Chen, J., Liu, C., Huang, R., Li, M., and Zhang, Y.: Determination of Geometrical and Structural Representative Volume Elements at the Baihetan Dam Site, Rock Mech. Rock Eng., 45, 409–419, https://doi.org/10.1007/s00603-011-0191-0, 2012.