R-09-38

Statistical methodology for discrete fracture model – including fracture size, orientation uncertainty together with intensity uncertainty and variability

Darcel C, Itasca Consultants SAS

Davy P, Le Goc R, de Dreuzy J.R, Bour O
Géosciences Rennes, UMR 6118 CNRS and
University of Rennes I

November 2009

Svensk Kärnbränslehantering AB
Swedish Nuclear Fuel
and Waste Management Co

Box 250, SE-101 24 Stockholm
Phone +46 8 459 84 00

SKB

ISSN 1402-3091
SKB Rapport R-09-38

Statistical methodology for discrete fracture model – including fracture size, orientation uncertainty together with intensity uncertainty and variability

Darcel C, Itasca Consultants SAS

Davy P, Le Goc R, de Dreuzy J.R, Bour O
Géosciences Rennes, UMR 6118 CNRS and
University of Rennes I

November 2009

Abstract

Investigations led for several years at Laxemar and Forsmark reveal the large heterogeneity of geological formations and associated fracturing. This project aims at reinforcing the statistical DFN modeling framework adapted to a site scale. This leads therefore to develop quantitative methods of characterization adapted to the nature of fracturing and data availability.

We start with the hypothesis that the maximum likelihood DFN model is a power-law model with a density term depending on orientations. This is supported both by literature and specifically here by former analyses of the SKB data. This assumption is nevertheless thoroughly tested by analyzing the fracture trace and lineament maps.

Fracture traces range roughly between 0.5 m and 10 m – ie the usual extension of the sample outcrops. Between the raw data and final data used to compute the fracture size distribution from which the size distribution model will arise, several steps are necessary, in order to correct data from finite-size, topographical and sampling effects. More precisely, a particular attention is paid to fracture segmentation status and fracture linkage consistent with the DFN model expected. The fracture scaling trend observed over both sites displays finally a shape parameter k_t close to 1.2 with a density term (α_{2d}) between 1.4 and 1.8. Only two outcrops (ASM000205 and ASM000206) clearly display a different trend with k_t close to 3 and a density term (α_{2d}) between 2 and 3.5.

The fracture lineaments spread over the range between 100 meters and a few kilometers. When compared with fracture trace maps, these datasets are already interpreted and the linkage process developed previously has not to be done. Except for the subregional lineament map from Forsmark, lineaments display a clear power-law trend with a shape parameter k_t equal to 3 and a density term between 2 and 4.5.

The apparent variation in scaling exponent, from the outcrop scale ($k_t = 1.2$) on one side, to the lineament scale ($k_t = 2$) on the other, addresses the issue of the nature of the transition. We develop a new "mechanistic" model that could help in modeling why and where this transition can occur. The transition between both regimes would occur for a fracture length of 1–10 m and even at a smaller scale for the few outcrops that follow the self-similar density model. A consequence for the disposal issue is that the model that is likely to apply in the "blind" scale window between 10–100 m is the self-similar model as it is defined for large-scale lineaments. The self-similar model, as it is measured for some outcrops and most lineament maps, is definitely worth being investigated as a reference for scales above 1–10 m.

In the rest of the report, we develop a methodology for incorporating uncertainty and variability into the DFN modeling. Fracturing properties arise from complex processes which produce an intrinsic variability; characterizing this variability as an admissible variation of model parameter or as the division of the site into subdomains with distinct DFN models is a critical point of the modeling effort. Moreover, the DFN model encompasses a part of uncertainty, due to data inherent uncertainties and sampling limits. Both effects must be quantified and incorporated into the DFN site model definition process. In that context, all available borehole data including recording of fracture intercept positions, pole orientation and relative uncertainties are used as the basis for the methodological development and further site model assessment. An elementary dataset contains a set of discrete fracture intercepts from which a parent orientation/density distribution can be computed. The elementary bricks of the site, from which these initial parent density distributions are computed, rely on the former Single Hole Interpretation division of the boreholes into sections whose local boundaries are expected to reflect – locally – geology and fracturing properties main characteristics. From that starting point we built Statistical Fracture Domains whose significance rely exclusively on fracturing statistics, not including explicitly the current Fracture Domains or closeness between one borehole section or the other.

Theoretical developments are proposed in order to incorporate the orientation uncertainty and the fracturing variability into a resulting parent distribution density uncertainty. When applied to both sites, it comes that variability prevails in front of uncertainty, thus validating the good level of data accuracy. Moreover, this allows to define a possible range of variation around the mean values of densities.

Finally a sorting algorithm is developed for providing, from the initial elementary bricks mentioned above, a division of a site into Statistical Fracture Domains whose internal variability is reduced.

The systematic comparison is based on the division of the datasets according to several densities referring to a division of the orientations into 13 subsets (pole zones). The first application of the methodology shows that some main trends can be defined for the orientation/density distributions throughout the site, which are combined with a high level of overlapping. Moreover the final Statistical Fracture Domain definition differ from the Fracture Domains existing at the site. The SFD are an objective comparison of statistical fracturing properties. Several perspectives are proposed in order to bridge the gap between constraints brought by a relevant statistical modeling and modeling specificities of the SKB sites and more generally conditions inherent to geological models.

Contents

1 Introduction

SKB has been leading geological investigations on the Laxemar and the Forsmark sites for several years. Much effort has been directed toward the sampling, understanding and modeling of the statistical properties of fracture networks inherent to the nature of the investigated rock. This in-depth investigation reveals the large heterogeneity of geological formations and of associated fracturing. But although the density of observations is exceptional, in particular with a detailed mapping of the fracture intensity all along the tenths of boreholes that were drilled in both Forsmark and Simpevarp, it still remains scarce compared to the natural heterogeneity of geological systems. For instance no detailed information about small-scale fracturing is possible in between boreholes, which leaves large volumes unexplored for fracture distribution. The basic multi-scale nature of fracturing is an additional challenging difficulty for integrating sparse information into site models. It requires a sound theoretical framework that makes possible the upscaling of local observations into a large-scale fracturing model.

In this project, we first justify the *DFN model framework* that will be used to assess fracturing parameters. We develop quantitative methods of characterization adapted to the nature of the fracturing process and to the datasets. Most of the methods that we present in the second part of this report are new and can lead to further improvement in future work.

The *DFN model framework* basically consists of the physical and mathematical expressions that eventually will be fitted to yield the site DFN model. The framework is built through a list of parameters (the *DFN model parameters*) that provides a quantitative description of the fracture network. The *DFN model framework* is defined by two sets of parameters: (1) the scaling parameters (power-law length exponent and fractal dimension), and (2) the density terms (coefficients defined by class of orientation, which are actual densities at a fracture length and domain scale of 1 m). In the following, the first parameters are called “scaling parameters”, and the second, “density parameters”.

In practice a given set of *DFN model parameters* is initially related to the spatial volume position and relative dataset from which it is derived. For instance, at depth, data arise from borehole logging. For the present analysis, the elementary datasets and volumes over which model parameters are computed are copied from the borehole local partition into Fracture Domains. The analysis is performed over the whole site and the initial division – into local limitations of Fracture Domains and consecutive initial datasets- is provided by SKB. The latter relies heavily on the local (single hole interpretation) lithological and apparent fracturing property consistency. DFN statistics are next performed independently from any additional consideration. It is deliberately a systematic comparison of fracture parameters which leads finally to the definition of *Statistical Fracture Domains* (SFD) over the site. No criterion is added to guarantee the spatial continuity of a SFD. Thus defined, the SFD are supposed to encompass volumes of similar fracture density. In this sense, they should be similar to the SKB *Fracture Domains* (FD) that are defined in the SICADA database. This correspondence is however further questioned in Chapter 5. Once defined, the SFD is intended to be used as a basic element of the *DFN site model* (see below) and, thus, as an elementary piece for interpolating between boreholes. The size of the fracture domains has thus to be consistent with this interpolation, which precludes defining domains much smaller than the average interpolation distance (approximately the distance between boreholes).

Finally, the *DFN site model* should be the ensemble of the *Statistical Fracture Domains* (each SFD is a *volume in space* with its corresponding *DFN model parameters* including their uncertainties). It is the end-product of the inverse problem that consists in retrieving the DFN model parameters from the database. The *DFN site model* is part of the general site model that contains additional information on geology, lithology, mineralogy, etc.

Many issues are raised by the DFN site model definition process. We list them below and refer to them throughout the report:

Issue 1 – The DFN model framework is chosen from different needs and requirements that can be contradictory: consistency with the physics of fracturing, derivation of the parameters from the existing database, and the **capacity for the DFN site model (see below) to bring constraints for further applications**.

Issue 2 – Fracture segmentation is a basic concern for DFN models for objective reasons: 1) although segmentation can be intrinsic to the fracturing process, like the en-echelon "Riedel" structures, it is more likely an artifact of sampling due to bad exposure quality of outcrops (ponds, grass, outcrop weathering); 2) even en-echelon or segmented structures can be connected below the surface exposure; 3) segmentation may dramatically modify the length distribution by artificially oversampling (undersampling, resp.) small (large, resp.) lengths; 3) segmentation generates highly correlated configurations (segment alignment) which are unlikely in a synthetic DFN generated by classical methods. The last reason, which is much looser, is that we would expect the *DFN model* to be a fundamental property of fracture systems that is valid over a broad range of scales. Power-law distributions were suggested to derive from self-organized processes, and fracture growth is cited as emblematic of that kind of processes /Turcotte and Rundle 2002/. By mixing up mechanical objects as different as complete fractures and fracture segments, the distribution may lose its universality.

Issue 3 – The "maximum-likelihood" *DFN model framework* is a **power-law length distribution with a density term that depends on orientation**. Definition of the scaling parameters (and, even more basic, the scaling function) is critical when predicting fracture densities at scales where no measure is available (roughly between 10 m and ca 400 m). The power-law scaling of fracture lengths is supported both by the literature /Bonnet et al. 2001/, and by the analysis of outcrop fracture traces at the different sites investigated by SKB. However, the connection between outcrop scale and lineament scales (~ 1–10 km) remains an issue when scaling exponents differ from one scale to the other.

Issue 4 – The definition of fracture domains proceeds from a combination of deterministic information, such as geological analysis that defines zones of similar "geology" (a general term that encompasses a series of petrological and deformation arguments), and from a systematic and quantitative analysis of fracture parameters. Moreover, the fracture density parameters in FDs cannot even be considered as uniform (or constant) because (1) Geology is never uniform on scales as large as those chosen for FDs (see Issue 5), (2) fracturing is a process that is intrinsically heterogeneous, and (3) the *DFN model* has an intrinsic statistical variability. In any case, characterizing this variability as an admissible variation of model parameter is a critical issue[1].

Note, additionally, that uncertainty is also attached to fracture parameters, and we do calculate it. **Variability** and **uncertainty** are fundamentally different in nature: **variability** means that fracture parameters vary, but we cannot say where; **uncertainty** means that fracture density may vary, but we cannot be sure it does. In stochastic simulations, variability can be treated as an uncertainty or, at best, as a stochastic function that ensures that fracture parameters vary through space.

Issue 5 – At depth, the only available information is from boreholes. Fracturing properties are defined by analyzing fracture intensities on cores, which represent a measure of the fracture density at the core diameter scale, i.e. ~ 80 mm. No information is available on the scaling parameters of the *DFN models*. These can only be assessed from outcrop fracture maps thanks to the exhaustive mapping of fracture networks over a sufficiently large scale range (Lineament maps pose a series of problems that make them useless for this assessment.). Then for FDs that have no clear correspondence with mapped outcrops (i.e. most of them) the scaling exponent cannot be assessed. But in any case, density parameters can always be compared in a relative manner, if a trial scaling exponent is fixed to a constant value everywhere.

Issue 6 – An objective when analyzing the along-core fracturing is to identify zones (core sections) of similar fracturing properties. The resulting fracturing classification is intended to be used as a starting point for sketching out the DFN site model. The accuracy of this process is intimately related to our ability to assert that two datasets lead to model parameters statistically equal or different. This statement is based on a comparison between the DFN parameters knowing both their uncertainty and their "intrinsic" variability within the volume considered. Two local datasets, or two FDs, are considered different if the difference between their DFN parameters is larger than the maximum of their uncertainty and variability. If not, and if variability is larger than uncertainty, we consider that they are similar. If not, and if uncertainty is larger than variability, we consider that they are *likely* to be similar but the reverse proposition may also be considered.

[1] The extrinsic variability due to geology factors can be reduced by subdividing fracture domain into small-scale pieces with uniform geological properties. However this is in contradiction with points (2) and (3), which require large domains to reduce the intrinsic variability.

Issue 7 about uncertainty, variability and the size of FD – The *DFN model* is basically a density function measured over a volume defined in space and size. Statistically the accuracy on the *DFN model parameters* follows the same rule as any averaging process: uncertainty decreases with the number of records. The determination of the volume over which the *DFN model parameters* is defined is thus a trade-off between uncertain knowledge of local values (on a small domain) and certain knowledge of non-local values (i.e. over a large domain). Variability may have a different behavior from uncertainty. Intrinsic variability is likely to decrease with system size for the same reasons as invoked above; but large domains may encompass lithological variations with a consequent increase of the internal variability.

The report essentially is organized in two parts.

The first part (Section 2) is dedicated to the assessment of fracture size distribution (Issue 2). This initially encompasses methodological developments for correcting outcrop data from observation biases and for providing best-estimators to power-laws. Both aspects impact the final exponents measured from the ten or so fracture trace-map datasets available from both sites. In this section, we propose a physical interpretation to link trace-map exponents and larger scales (Section 2.6).

The second part of the report focuses thoroughly on variability and uncertainty modeling (Section 4) coupled to the *statistical fracture domain* assessment (Section 5), or, equivalently, to the site-scale modeling (*Issues 3* to *6*). The guidelines for this part are seen in Figure 1-1. The method eventually aims at finding an optimized partitioning of the site into relevant *statistical fracture domains*, given the overall uncertainty, variability and data availability conditions. For that purpose, we initially developed the modeling framework to handle both variability and uncertainty (orange in Figure 1-1). Then we developed an automated sorting algorithm (blue in Figure 1-1) in order to group dataset density distributions into "homogeneous" classes based on internal variability reduction. This part requires proper definition of the notion of "compatibility" between density distributions. Also, the quantitative method allows us to embrace simultaneously any number of datasets. It combines the three facets (variability, uncertainty and determinism) throughout the initial subdivision of the fracture database, the variability and uncertainty assessment, and the final sorting of the DFN models into compatible classes.

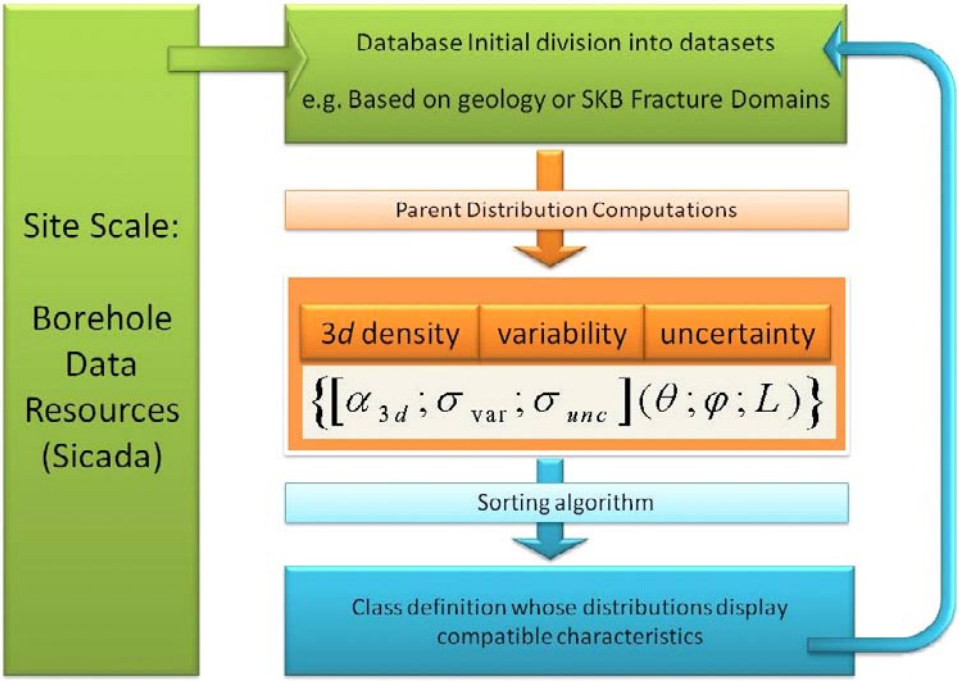


Figure 1-1. *Overview on the DFN modeling at the site scale, including the variability and the uncertainty, applied to the borehole database.*

Note that this analysis was made possible because:

- Borehole data are numerous and well distributed all over the site, and presumably, sampling various fracturing conditions.
- Fracture-pole intercept orientations are associated to measure uncertainties.
- Sampled borehole sections are generally sufficiently large to allow assessment of the variability scaling.

Based on geological interpretations and modeling already performed by SKB, only the fracture-pole intercepts belonging to FDs [in opposition to Deformation Zones (DZ)] are included in the analyses (Section 3). The complete analysis requires calculating tens to hundreds of parameters from the SICADA database. A significant effort has been put into developing semi-automatic routines to select data and calculate parent density distributions with uncertainty and variability on parameters.

2 Fracture sizes: Scaling Model

2.1 Introduction

This chapter focuses on the analysis of fracture size distributions. All available data from both Laxemar and Forsmark sites are considered, from fracture trace maps up to lineament and Deformation Zone maps. It covers a broad range of scales. In order to highlight the multiscale nature of fracture sizes, analyses of the fracture size distributions are presented into log-log diagrams. Indeed, the power-law scaling of fracture traces is supported both by the literature /Bonnet et al. 2001/ among others) and by previous analyses performed at the outcrop scale towards SKB sites. Before entailing the interpretation stage, a strong attention is paid on how to derive appropriate distribution parameters from field data.

Several corrections are indeed necessary to go from initial raw data to final size distributions, from whom DFN model parameters are derived. Essentially this concerns two aspects: sampling biases and adequacy between fracture model definition and apparent data (see Issue 2 in Introduction). Common biases associated to fracture data sampling are due to **truncation** and **censoring** effects /Pickering et al. 1995/. The main cause of censoring is that fracture traces may cross-cut the outcrop boundaries, thus preventing from defining the trace length. Truncation occurs since the size of the sampling window is limited – bounded below by the resolution of the outcrop mapping and above by the scale of the outcrop – and, in practice, significantly smaller than the range of sizes covered by the fracture size distribution. In addition, significant biases can arise from **apparent segmentation** and deformation of fracture traces, due to exposure conditions, weathering and map departure from planarity (local topography). All these factors directly influence the distribution shape (Figure 2-1). In addition, the level of interpretation required to interpret data according to the fracture model meaning – for instance when several en-echelon segmented traces are merged into one feature- will also impact the distribution shape. This disconnection leads to an overestimate of small fractures at the expense of large fractures.

Several existing techniques for correcting classical sampling bias can be cited: /Priest and Hudson 1981/ devised formulae to identify the underlying distribution from a biased sample, assuming an exponential parent size distribution. Later, /Munier 1995/ applied this method to power-law distributions. Because, in some circumstances, these methods can assume non standard forms for which there are no standard analytical tools, they might be conceived as being counterintuitive. Instead of correcting the fracture size distributions, other approaches mimic the real sampling conditions (such as alteration, map shape). An objective function is defined to compare artificial fracture traces and field observations /Dershowitz et al. 1991a, b/. These methods are based on stochastic simulations associated to simulated annealing or a conjugated gradient algorithm.

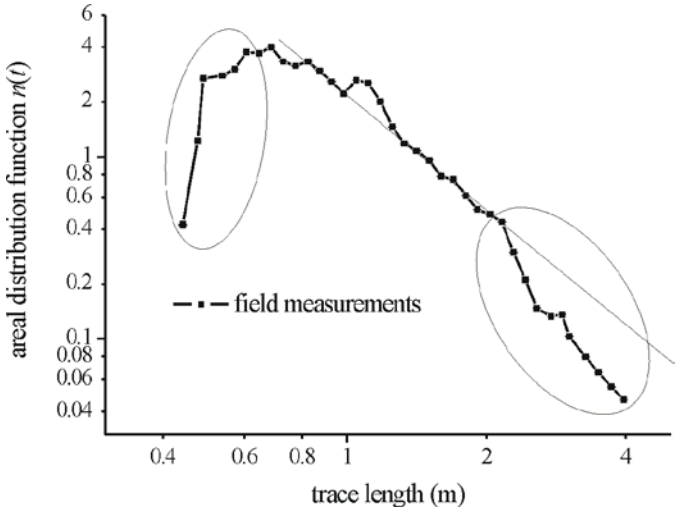


Figure 2-1. *Fracture trace length distribution calculated from a raw outcrop trace dataset: the power-law behavior of the distribution is visible in the middle part. Right-hand and left-hand sides of the distributions are affected by censoring, truncation and segmentation effects.*

We rather start the correcting process by a treatment focused on each fracture trace according to its position into the rest of the network; this encompasses first a topographical correction presented in section 2.4.1, followed by the main point, the linkage process between some fracture traces (presented in section 2.4.2). Next the finite size correction are applied (description in section 2.4.3) and the fracture trace length distributions can be interpreted. We notice that the fracture trace size distribution parameters estimation is very sensitive to noise /Oltmans and Verheijen 1997/. We therefore consider initially two systematic fitting processes (least-square and maximum likelihood) and select the best adapted to power-law parameter estimation (section 2.3). Results and interpretations on the fracture size model relevant at various scales of the analyzed sites are provided in section 2.5.

2.2 Observation scales and data resources

The various observation scales available in the database are illustrated in Figure 2-2. Several main range of sizes are defined: (a) the **outcrop** or **fracture trace map data,** from less than one meter and up to several meters (green in the figure), the (b) **lineament** and (c) **deformation zone** maps, from a hundred of meters and up to few kilometers (black, red and navy blue in the figure) and finally, (d) the country scale map, made of lineaments covering all Sweden, from 10 km and above (orange in the figure). This latter map does not belong to the official SKB database. However, it provides a very large scale trend, which is moreover common to both sites**.**

The data collection-sampling process and required interpretations for defining fracture trace tip positions obviously varies for the various maps and scales mentioned. The outcrop map sampling is very similar from one outcrop to the other and from one site to the other. Observations are direct and few interpretations are performed before inserting these observations in the database. Specifically, fracture segmentation is not addressed. Further interpretations/corrections are therefore necessary before deriving the fracture size distribution from the discrete data. This is described in sections 2.3 and 2.4 in the present chapter.

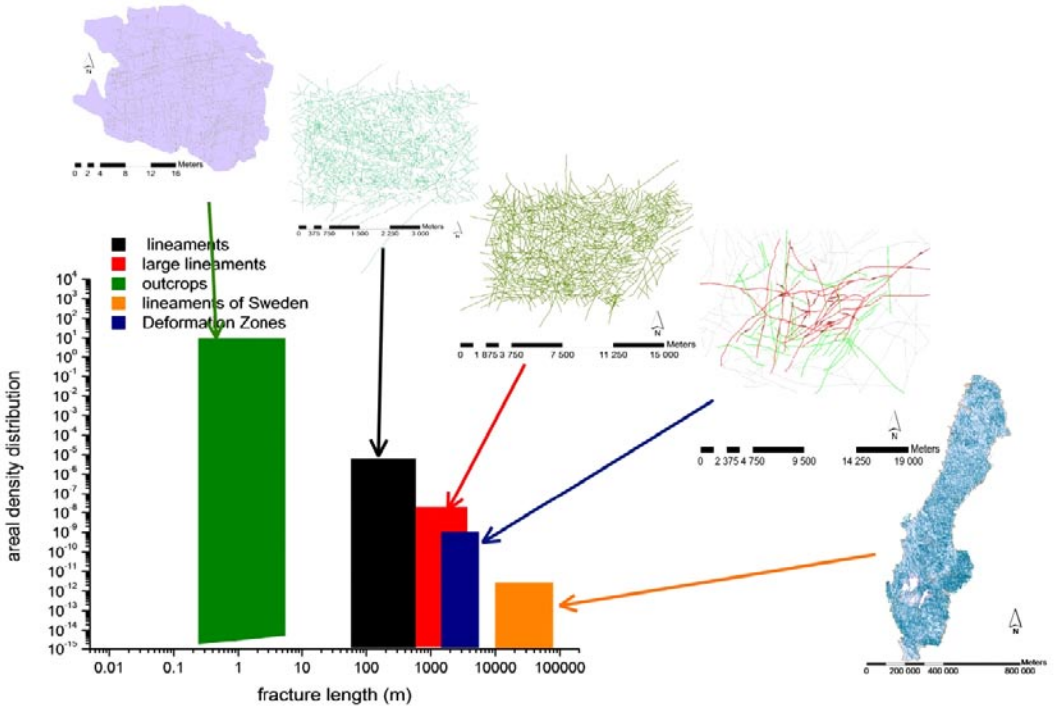


Figure 2-2. *Typical range of scales into which fractures are sampled. Fracture trace maps are symbolized in green, lineaments in black and red and deterministic Deformation Zones in navy blue (illustrating maps correspond to Laxemar site). The largest range of scales defined , in orange, refers to the Sweden lineament scale, from 10 km and above.*

At larger scales, direct observations are not possible and one needs to resort to interpretations in order to infer discrete sets of lineament traces from several geophysical data collections (for instance ground magnetic, topographical analysis, seismic…). Moreover, a ultimate step in the interpretation process is necessary to identify DZ among the lineaments. **The data delivered already encompass these stages**. Figure 2-3 below illustrates further how the "lineament" and DZ maps are embedded. For both sites two lineament maps are present, one within roughly a 2–3 km side area (dark grey traces in Figure 2-3) and the other at the regional scale (10 km side box). Moreover, the DZ map in Laxemar is even larger than the regional scale whereas is it exactly the same for Forsmark. Knowing these discrepancies, all the above-cited "lineament" maps are considered in the following analyses. We thus refer to the lineament maps as "sub-regional" (dark grey in the figure) and as "regional" (green in the figure).

We recall finally that the delivered lineament datasets are already co-ordinated or linked, such that applying the reconnection procedure developed for the outcrop trace maps is not appropriate. Only the finite size corrections are similarly applied on all the datasets.

Most of the data used in the scaling analysis come from the SICADA delivery #SICADA_08_32 (section Data resources for the fracture size analyses). Two additional datasets complement the main delivery: the Forsmark regional scale model (*SDEADM_GV_FM_GEO_2079*) and the Sweden scale dataset (mentioned above).

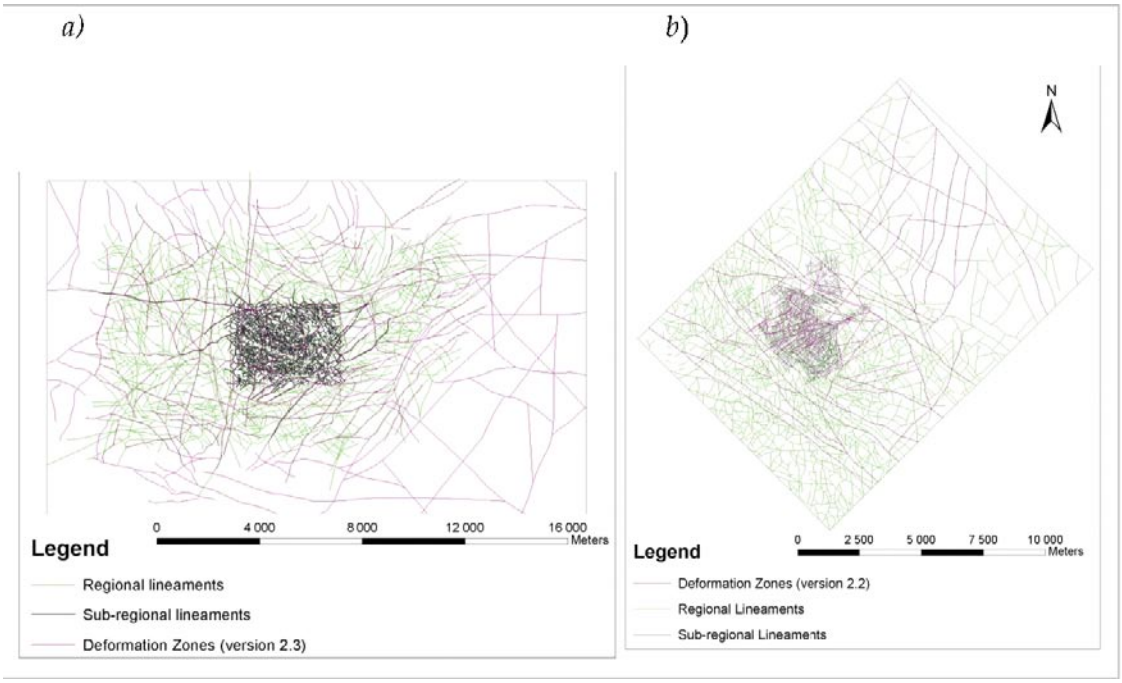


Figure 2-3. *Surface map of lineaments and Deformation Zones, scale overlapping (complementary views of these maps are provided in Figure 9-1).* ***a)*** *Laxemar and* ***b)*** *Forsmark.*

2.3 Fitting method

A complete description of the power-law model can be found in /Newman 2005/. The standard strategy to determine power-law behavior is to create a histogram of the distribution, which appears as a straight line when plotted on log-log scales. The probability density function (*pdf*) of a power-law distribution can be generally expressed as:

$$f(x) = \frac{k \cdot x_{\min}^{k}}{x^{k+1}} \qquad 2\text{-}1$$

Where x_{min} is the smallest value of x and $-(k+1)$ the scaling exponent. In the DFN modeling framework, when the distribution refers to a fracture diameter distribution, then x_{min} is equivalent to the location parameter r_0 and k is called the shape parameter as defined in /Munier 2004/. In complement, the fracture trace distribution model, noted n (.), encompasses the density term α and is related to Equ. 2-1 by:

$$n(x) = \alpha \frac{x_{\min}^{-k}}{k} f(x) \qquad 2\text{-}2$$

The distribution parameters can be visually fitted from the plot of the histogram distribution on a log-log diagram. However the manual fitting is prone to errors and subjective biases, especially on the right-hand side of the distribution where the power-law distribution dwindles (each bin/point in this region has a weak statistical weight, i.e. very few data). In order to avoid potential biases due to manual fitting, we consider the use of a systematic fitting, based on least-squares procedure or maximum-likelihood estimator methods.

The **least-squares** fit method is based on the square differences between true values y_i and modeled values $y_{mod i}$ of the distribution. Each point has a standard deviation, σ_i , so that we can compute the χ^2 value through

$$\chi^2 \equiv \sum_{i=1}^{M} \left(\frac{y_i - y_{\mathrm{mod}\,i}}{\sigma_i} \right)^2 \qquad 2\text{-}3$$

with M the number of points in the distribution. The model parameters are obtained by minimizing χ^2.

The second method is based on maximum likelihood estimators /Press et al. 2007/. The maximum likelihood estimate relies on the maximization of the probability that the observed distribution comes from the modelled one. The method leads to the following estimation of k /Newman 2005/:

$$k = N \left[\sum_{i=1}^{N} \ln \left(\frac{x_i}{x_{\min}} \right) \right]^{-1} \qquad 2\text{-}4$$

where x_{min} is the minimum value considered and N is the number of elements (in practice here fracture traces) in the dataset. Following Equ. 2-1 and Equ. 2-2 the density term is:

$$\alpha = \frac{N}{A} \cdot k \cdot x_{\min}^{k} \qquad 2\text{-}5$$

with A a normalization factor – surface of a trace map for the present application.

Estimates of the expected statistical error on k and α, resp. σ_k and σ_α are given by

$$\sigma_k = \sqrt{N} \left[\sum_{i=1}^{N} \ln \frac{x_i}{x_{\min}} \right]^{-1} = \frac{k}{\sqrt{N}} \qquad 2\text{-}6$$

and

$$\sigma_\alpha = \frac{N}{A} \cdot x_{\min}^{k} \cdot \left(\sigma_k + k \cdot \ln \left(x_{\min} \right) \right) \qquad 2\text{-}7$$

A comparison between the two fitting methods is displayed in Figure 2-4. Firstly, the maximum likelihood estimation is less sensitive to outliers (Figure 2-4*a*). Secondly, the largest scale behavior seems to be better catched by the maximum likelihood rather than by the least square fitting. The resulting differences in the founded exponents can be quite important. We therefore chose to use the maximum likelihood estimator in the analyses.

2.4 Applied corrections

2.4.1 Topography

The first step of the correction procedure is to reduce bias caused by the topography. Maps defined for fracture trace sampling are not perfectly planar but rather naturally irregular. This potentially entails a bias into the recording, such as apparent fracture trace segmentation due to sharp accident or trace curvature increase due to round topographical variation. This effects are naturally more pronounced for subhorizontal fractures. They can affect the following fracture trace linkage process.

The basic idea is to project all fracture traces on a flat surface, in order to reduce the perturbation due to topography. We choose as reference plane $\Pi: z = z_{mean}$, where z_{mean} is the mean elevation of the fracture trace coordinates. Note that the GIS database contains only the 2D, or North-East coordinates, initially recorded positions of these traces. The z-coordinates, or elevation, of all traces are nevertheless available in anterior files (specific delivering references #SICADA_07_042). Fractures are modeled by flat planes, whose orientations are the strike and dip. Each trace segment is projected along the fracture plane in the fracture dip direction (Figure 2-5). The directing vector of the projection can be written.

$$\vec{v} = \begin{pmatrix} -\sin(\theta)\cdot\cos(\varphi) \\ \cos(\theta)\cdot\cos(\varphi) \\ -\sin(\varphi) \end{pmatrix}$$

Where θ is the fracture strike and φ is the fracture dip.

In Figure 2-6*a* and *b*, an example of non-projected and projected traces is provided. The projection tends to straighten curved traces for fractures of small dip. Figure 2-6*c* shows the lack of influence of the projection on the final fracture trace size distribution. Fractures shortened through the projection do not significantly influence the estimate of the fracture length distribution. They essentially fall within the range of truncation effects.

In general, for the fracture trace maps analyzed, the topography entails almost no direct impact on the size distribution shape. However, it has a more significant role in fracture traces linkage.

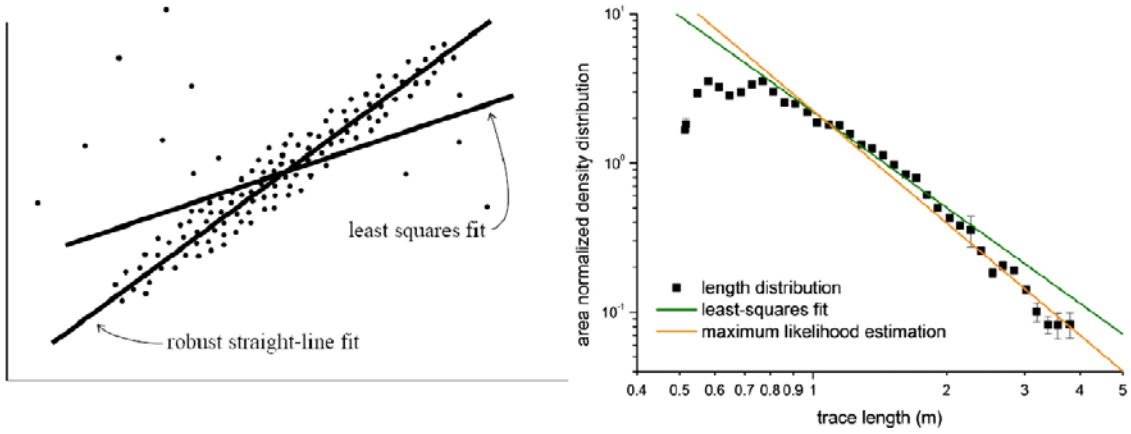


Figure 2-4. *Comparison between least squares and maximum likelihood estimators.* ***a****) Typical example of fitting by maximum likelihood and least-square procedures in the presence of outliers (from /Press et al 1992/).* ***b****) Application to a fracture trace size distribution; the least-square fit (green) does not catch the most right-hand side of the distribution, whereas the maximum-likelihood estimation (orange) does.*

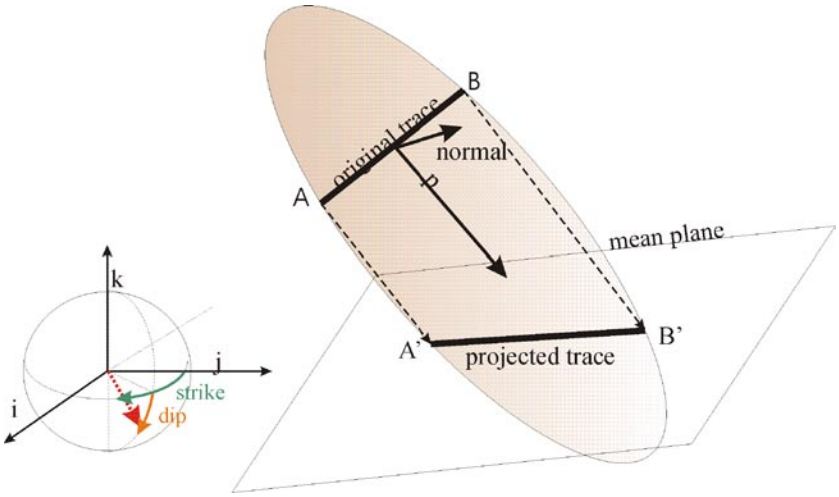


Figure 2-5. *Illustration of the projection of the trace coordinates to a mean plane along the fracture direction.*

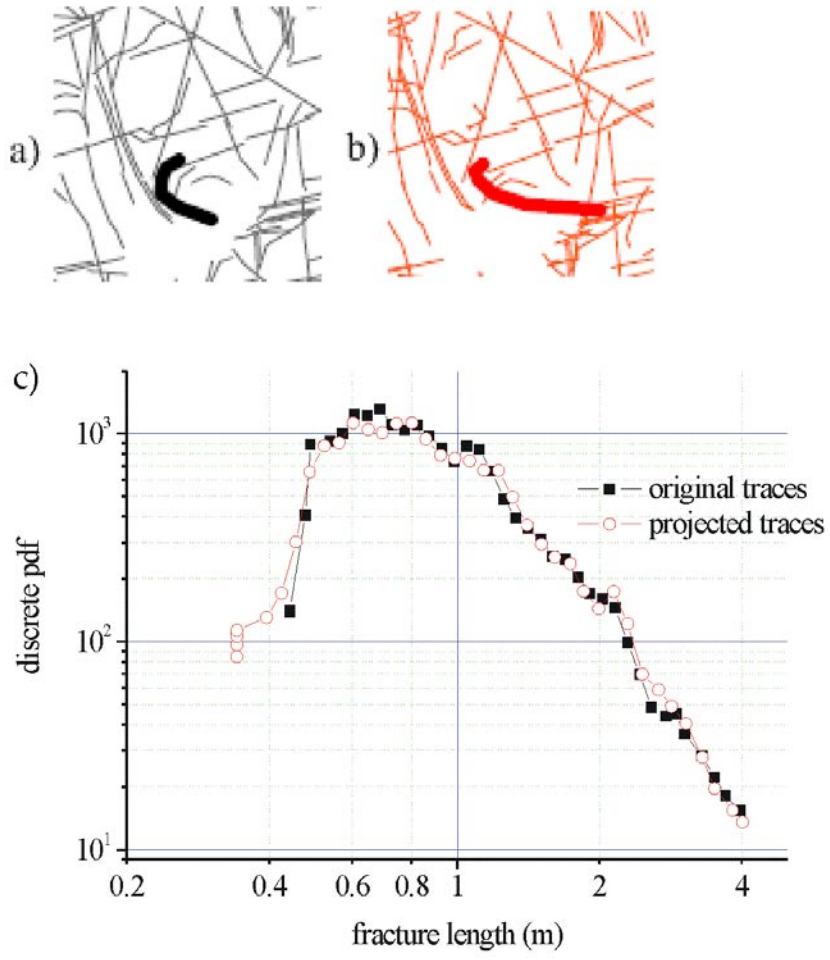


Figure 2-6. *Example of a non projected a), and projected b) trace (trace map ASM000208 from Laxemar). The fracture is associated to a small dip value. The corrected trace is straightened. c) Fracture trace length distribution for the original data (black, squares) and for the projected data (red, circle). No crucial influence of the projection in the power-law parameters is noticed.*

2.4.2 Linkage of segmented traces

The second step concerns the identification, and consecutive linkage, of apparently segmented fracture traces (Figure 2-7). As mentioned in Introduction (issue 2), these configurations reflect in general sampling limitations due to weathering, or an intrinsic outcome of the fracturing process, like the en-echelon "Riedel" structures.

For fracture trace pairs whose respective pole orientations are close enough, the level of overlapping reflects all the range of possible segmented trace configurations. This is simply computed by considering the orthogonal projection of each tip of the trace on the other trace forming the selected pair. For an increasing level of overlapping, one encounters initially some dashed-line configurations (Figure 2-7*a*), then slight overlapping, then en-echelon or "Riedel" structures when overlapping is large enough, and finally "layered configurations" when overlapping reached 100% (Figure 2-7*c*). The latter configuration occurs in less than 1% of the cases. In such cases no merging occurs and the traces are left unchanged. In summary fracture traces to be linked should be recognized according to two intuitive criteria: similarity of their respective orientations (solid angle between fracture poles) and vicinity of their tips (Figure 2-7).

The orientation criterion is simply based on the solid angle between fractures, which must not exceed a threshold (noted ω_{max}). In addition, a critical area around fracture tips is defined for selecting potential configurations**.** This latter derives from mechanical considerations. A zone of influence is defined around each fracture tip (grey dashed line elliptic shapes in Figure 2-8). Links can be made only in this zone. For each fracture tip the zone of influence is an ellipse, defined in polar coordinates. Its size is related in principle to the stress conditions and to the fracture length (t_{frac}) itself, such as /Atkinson 1989/:

$$\frac{\Delta\sigma_I}{\tau} = \sqrt{\frac{t_{frac}^{\ 2}}{4 \cdot r^2} \times f(\theta)} \qquad 2\text{-}8$$

where $(r;\theta)$ are the polar coordinates of the observation point (origin at the fracture tip), τ is the shear stress induced by the fracture at the fracture tip /Atkinson 1989/, according to:

$$\tau = \frac{1}{2}\sqrt{(\sigma_{22} - \sigma_{11})^2 + 4\sigma_{12}^2}$$

with σ_{11}, σ_{22} and σ_{12} the components of the stress tensor, $\Delta\sigma_I = (\sigma^r_{11} - \sigma^c_{11})$ the difference between remote stress, σ^r_{11}, and the stress on the crack, σ^c_{11}.

One fracture trace is selected for linkage if it intersects the curve given by Equ. 2-8. In the linkage process, this defines the second criterion: a distance threshold $(\tau/\Delta\sigma_I)_{max}$ which reflects the stress conditions. The zone of influence increases with increasing $(\tau/\Delta\sigma_I)_{max}$. Note that the actual threshold value will be based on statistical considerations (Figure 2-10).

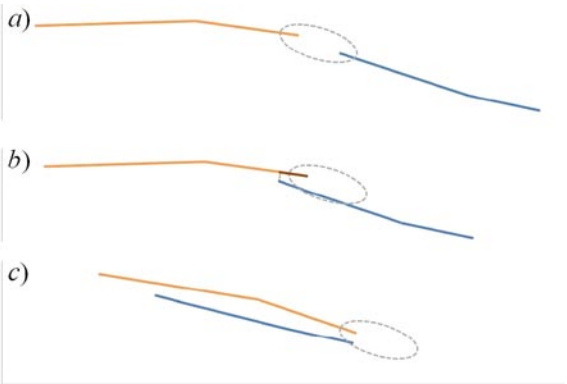


Figure 2-7. *Example of non-Poisson structures in the natural fractures.*

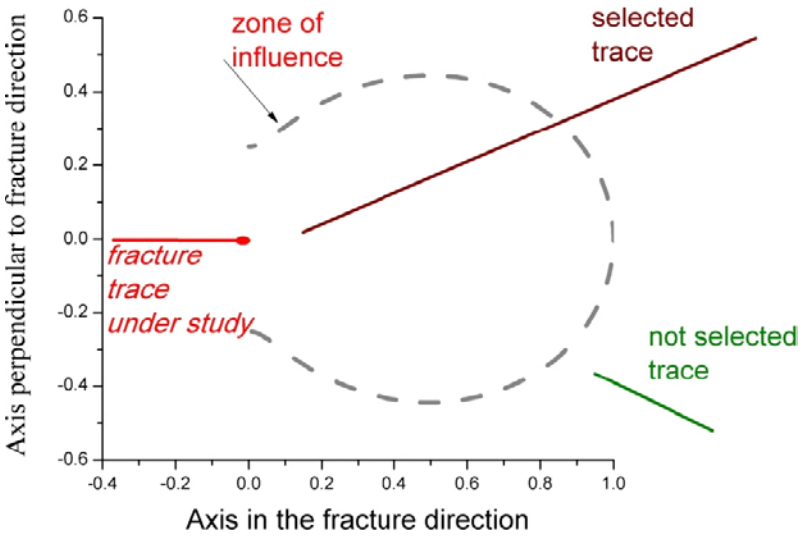


Figure 2-8. *Contour of shear stress normalized with fracture length.*

Based on the two criteria thus defined the whole linkage process can be performed. The linkage algorithm is summarized in Table 2-1 below.

Figure 2-9 finally illustrates how the reconnection is performed. When there is no overlapping between traces, a new segment is added to join the tips (Figure 2-9*a*). When overlapping is not null, the segment connecting the projected points is added while the fracture tips are accordingly adjusted (Figure 2-9*b*).

The critical point of the linkage process consists in finding the most appropriate threshold values for the orientation and distance criteria. One obvious test consists in visually checking a fracture trace map after reconnection (step 12 in Table 2-1): we check that no awkward trace appears: the linkage segment added must not change the general shape of the parent pair of fracture traces.

Table 2-1. Fracture trace data corrections and linkage algorithm (between dot lines).

(1) *Projection of fracture traces on a flat topography.*

---

(2) Set orientation and distance criteria, resp. ω_{max} and $(\tau/\Delta\sigma_I)_{max}$.

(3) For all pairs of fracture traces do:

- **(4)** Compute the ellipse near the larger fracture tip and the orientation between traces.
- **(5)** If distance and orientation criteria are fulfilled and the traces do not intersect.
  - **(6)** Save the current pair as a possible reconnection.

(7) For all possible reconnections:

- **(8)** In case of competition between pairs (involving the same fracture), select only the pair with the nearest fracture.
- **(9)** Apply the reconnection.

(10) Return to step (3) if the number of effective reconnections was larger than zero.

(11) Check visually the resulting fracture network:

- **(12)** If the fracture network is consistent (no unrealistic trace) increase ω_{max} or $(\tau/\Delta\sigma_I)_{max}$ and return to step (2) or go to step (13), else decrease ω_{max} or $(\tau/\Delta\sigma_I)_{max}$ and return to step (2).

---

(13) *Correct the finite-size effects.*

(14) *Compute the trace-length distribution parameters.*

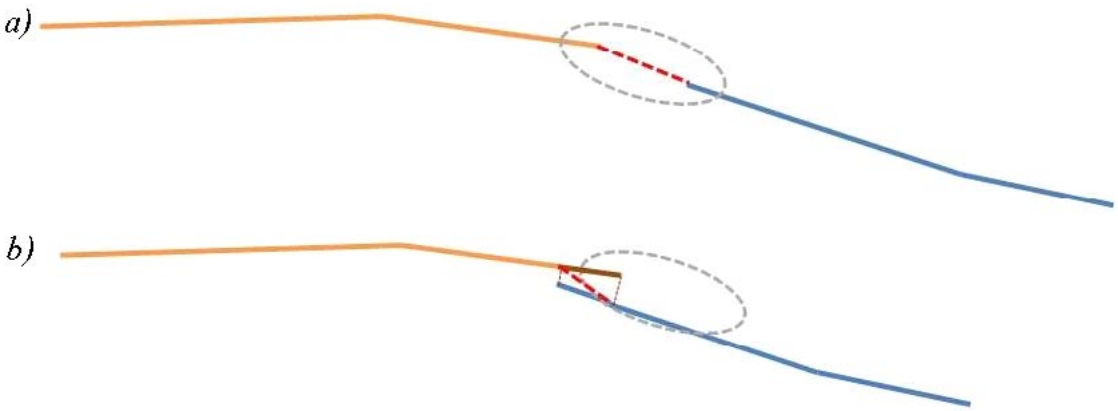


Figure 2-9. *Description of reconnection procedures.*

Moreover, the **orientation criterion** is chosen in order to reflect the orientation variations between the fracture trace and its constitutive segments. For each trace map, an angle distribution is accordingly computed, whom mean and standard deviation are added to define a critical angle. For the Laxemar outcrops the values fall within 20° and 25°. We retain finally a value – ω_{max} = 20° – which minimizes potentially the number of abnormal reconnections.

Finally, we chose to stop the linkage process at a level above which increasing linkage would lead to observe less segmented configurations than in an equivalent artificial network, where spatial correlations between fracture trace length and position are removed. For doing this we consider a comparison between real fracture trace maps and artificial networks where fracture traces are purely randomly spatially redistributed within the outcrop map (the initial orientation and trace length distributions are kept). For each outcrop map dataset, several realizations of the thus defined artificial networks are generated.

The amount of segmented structures in a network is necessarily given in relation with the distance criterion used for the identification process. We thus compute the ratio $N_{seg}(d_{max})/N_{tot}$ given by:

$$\frac{N_{seg}(d_{\max})}{N_{tot}} = \frac{\text{Number of segmented structures according to d}_{\max}}{\text{Number of fracture traces}}$$

The evolution of this ratio is represented for artificial (grey lines) and real fracture trace maps reconnected at various stages (colored symbol plus lines) in the following Figure 2-10.

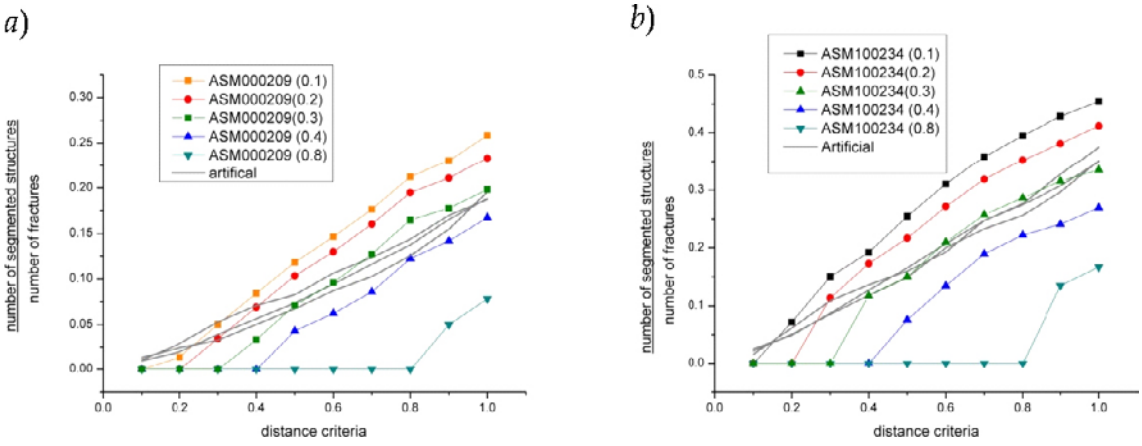


Figure 2-10. *Evolution of the amount of segmented structures as a function of the distance criteria. a) dataset ASM000209 and b) dataset ASM100234.*

For the artificial networks, 3 realizations are displayed, where the number of apparent segmented structures is computed from various distance criteria. From one realization to the other we observe a similar behavior. For the initial natural datasets, we consider various levels of linkage (see inlet in Figure 2-10) and the same range of distance criteria (x-axis in Figure 2-10). Accordingly, the ratio $N_{seg}(d_{max})/N_{tot}$ is null for a network already reconnected up to a distance d_{max} and it increases above. We then look for the optimal distance criterium that produces next a number of segmented structures equivalent to what arises from a non-correlated process (i.e. the artificial network). In the present case, the optimal distance criterion ranges between 0.3 and 0.4. The linkage process will not be pursued further. By doing this, we replace spatial correlations reflected in apparently segmented structures by single linked traces, so that the final trace length distribution directly encompasses these correlations. Also, the process is limited in a way such that after linkage the highest sources of spatial correlations, the segmented structures, are merged. An illustration of the linkage process is presented in Figure 2-11.

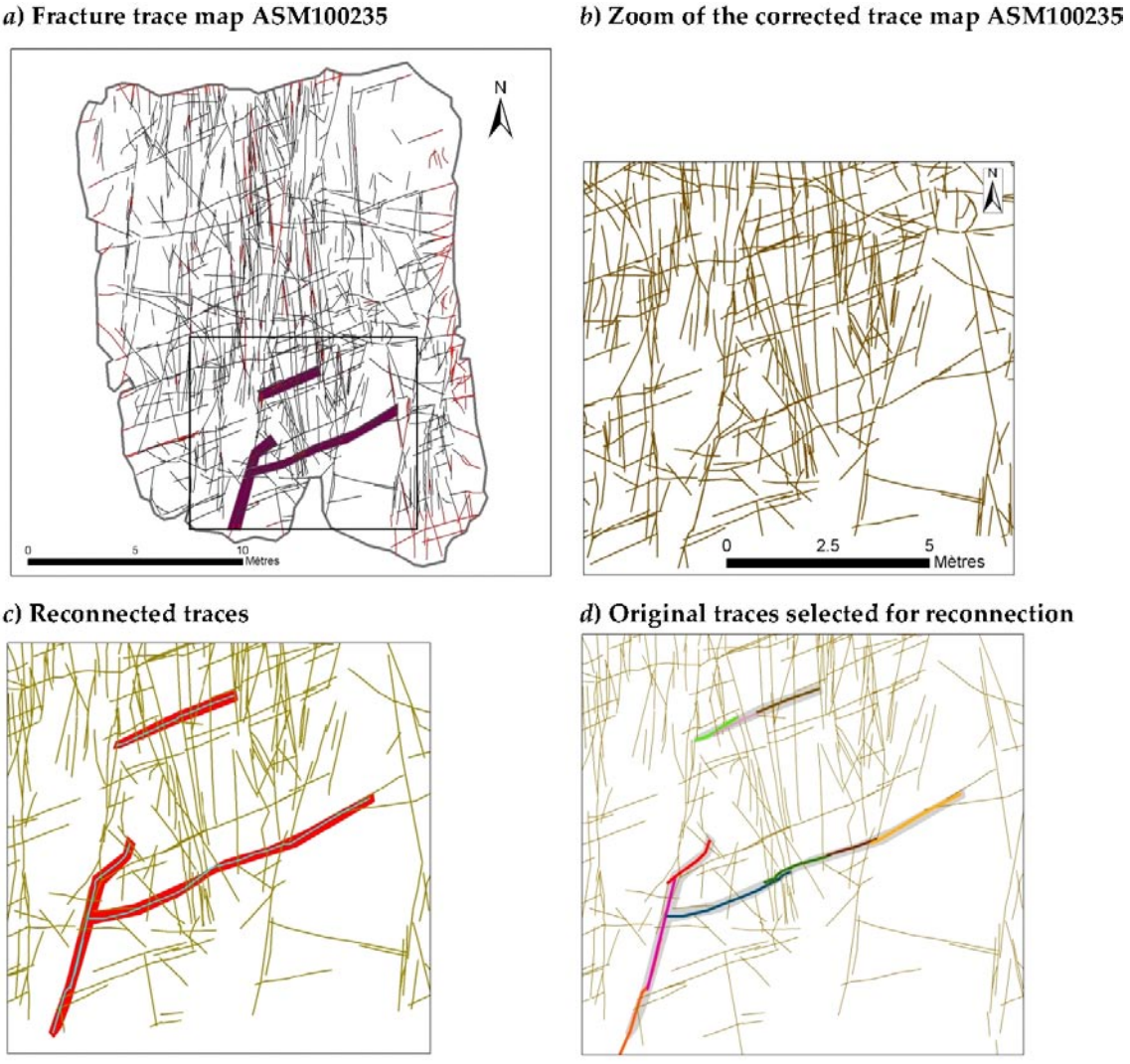


Figure 2-11. *Illustration of linkage on the fracture traces from outcrop ASM100235. a) The entire network, b) detail of the reconnected network. c) highlighting, in red, of 3 reconnected traces and d) corresponding original segmented traces.*

2.4.3 Finite size correction

In an outcrop, the probability of seeing entirely a fracture trace is related to the fracture trace length itself. This is taken into account for the fracture trace length distribution analysis: each point of the distribution (from the number of traces whose length is between t and $t+dt$) is computed over an area where the corresponding fracture trace can be entirely observed. This comes to define the area in which the fracture center can be located. For each range of size considered (each point of the density distribution) the normalization area is accordingly adjusted.

In practice, for each range of fracture trace length, the right area is computed by using the projection of a regular mesh, whose edge cell is roughly one hundredth of the outcrop side (Figure 2-12). We count only cells at a distance greater or equal than $t/2$ from the boundary, where t is the mean fracture-trace length of the current class, and multiply this number by the area of a cell (Figure 2-12).

In summary, each point of the fracture trace length distribution arises from traces entirely included in the sampling surface. The distribution is normalized by the corrected area described above.

2.5 Results

In the following, we present the resulting exponents of the power-law distributions after correction at all available scales. We recall that although fracture trace and lineament datasets allow us to derive size distributions, they differ slightly in nature and type: trace data are not interpreted, and encompass precise fracture position, trace length, dip and strike. Lineament data, however, already are interpreted (linkage from co-ordination between geophysical methods) and do not contain the dip information.

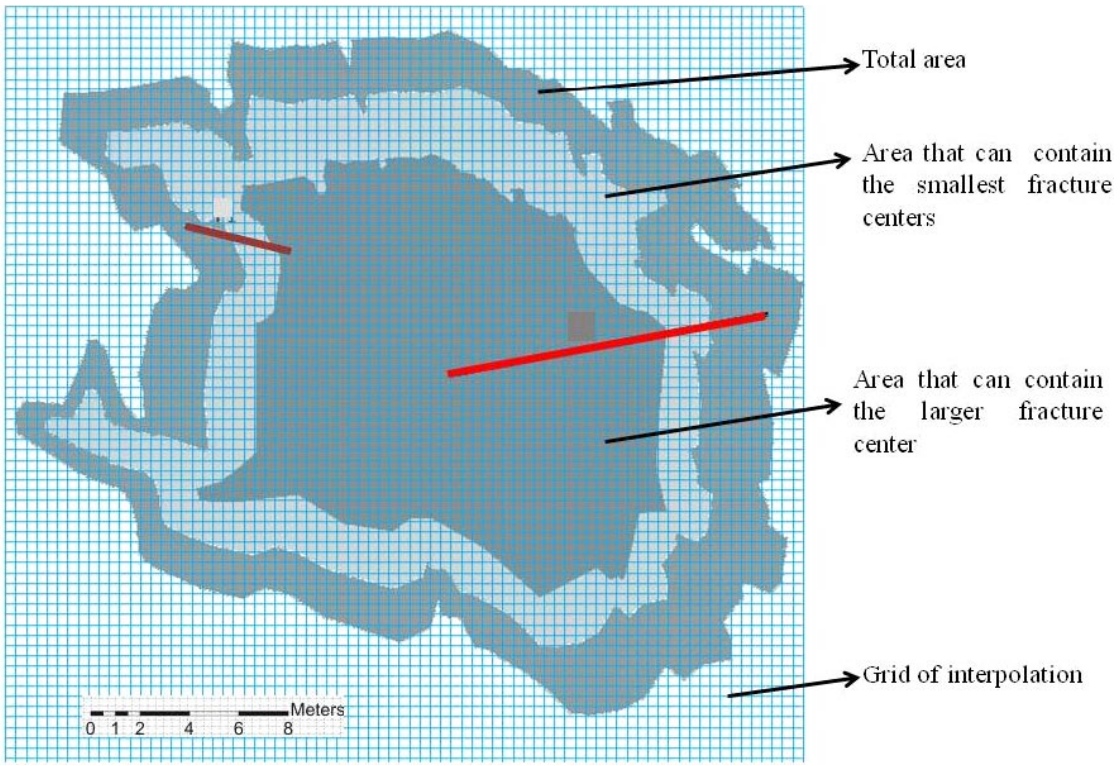


Figure 2-12. *Representation of area normalization accounting for the finite-size effect. The range of large fractures (symbolized by a red trace)is normalized by an area smaller than the one attributed to smaller fractures (symbolized by a brown trace). The difference between these areas causes a bias in the observations.*

2.5.1 Impact of the corrections

At the outcrop scale, the linkage process leads to stretch out the traces. As illustrated in Figure 2-13, this has significant impact on the global shape of the fracture trace size distribution. Indeed, the fracture trace size distribution based on a corrected dataset (including the finite size correction) displays a reinforced straight trend (i.e. over a larger range) when compared to the non-corrected one.

2.5.2 Scaling exponents from trace-map scale

A compilation of all measured fracture-size distribution parameters is provided in Figure 2-14 and Figure 2-15. The first displays only power-law exponents on separate diagrams according to site reference; the second displays (α_{2d}) = f (k_t+1), the density term vs. the power-law exponent evolution, for the raw and corrected data. "Global correction" refers to the linkage including topography and to the finite size corrections.

As highlighted in Figure 2-14, a global correction entails a significant decrease of the power-law exponent (k_t). For instance, for the fracture trace maps at Laxemar, the initial mean without correction, k_t , is equal to 1.65 , whereas it decreases to 1.21 after correction. The scaling-exponent decrease is related logically to the size increase of some fractures during the reconnection process. Therefore, due to apparent disconnection, the original data appear to under-rate large fractures and over-rate small fractures.

Thus, the type of correction applied entails significant variations of the resulting parameters of the fracture size distribution model. In complement, further observations reinforce the confidence in the corrections.

- The range of exponent values is tightened after correction. In the DFN modeling perspective, we define the mean power-law exponent and dispersion by averaging the values derived from each dataset of a site (disks and error bars in Figure 2-14). It appears that dispersion around the mean decreases when the whole correction process is completed. At Forsmark, the square difference between the different exponents has a value of 0.21 without correction and a value of 0.10 after correction.
- The raw data synthesis diagram (Figure 2-15*a*) displays an apparent correlation between both parameters of the power-law model (dashed grey line in the figure). It follows the trend:

$$\alpha_{2d} \sim 3^{(k_t+1)}$$

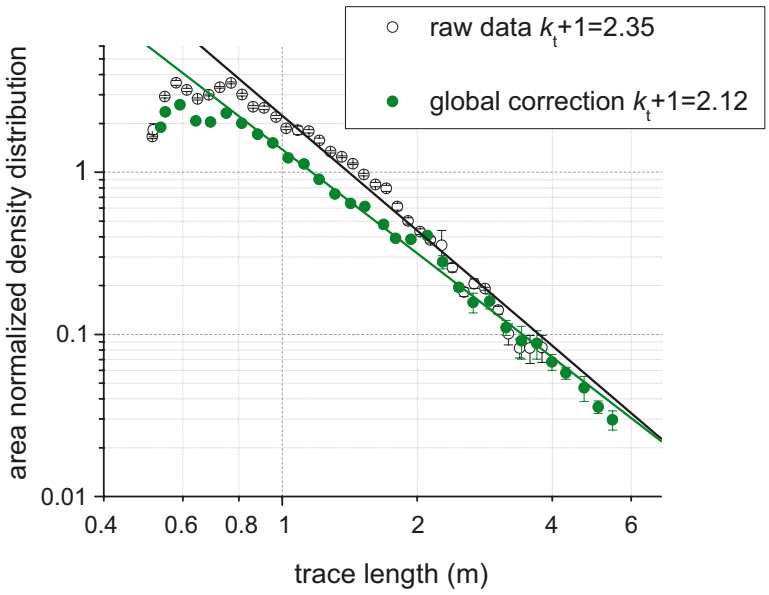
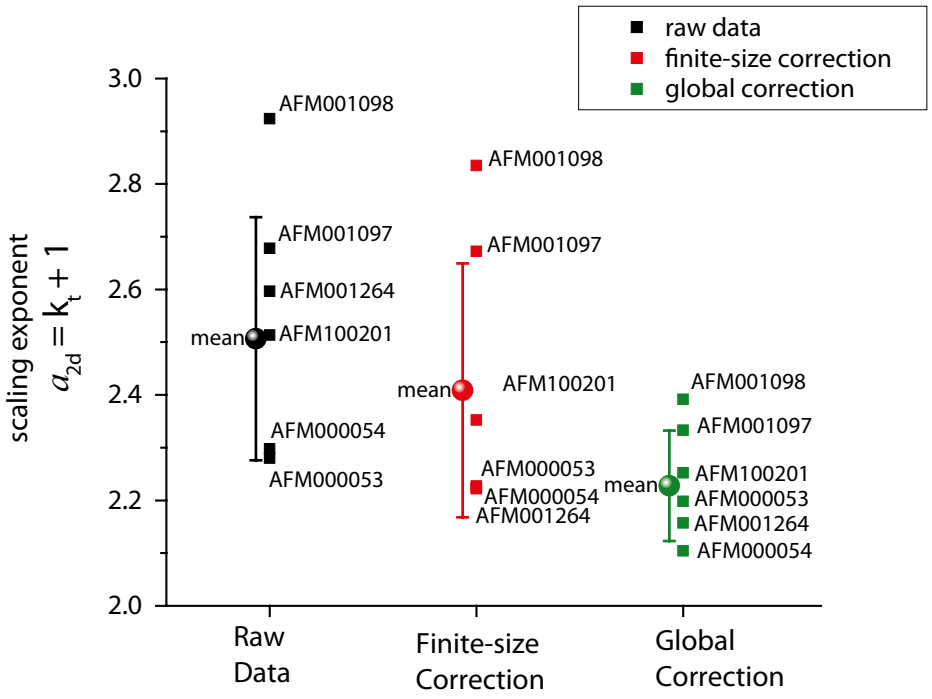


Figure 2-13. *Fracture trace size distributions from outcrop ASM100234 (Laxemar): from raw data in black and empty circle, and after corrections (topography, linkage and finite size) in green circles.*

a)



b)

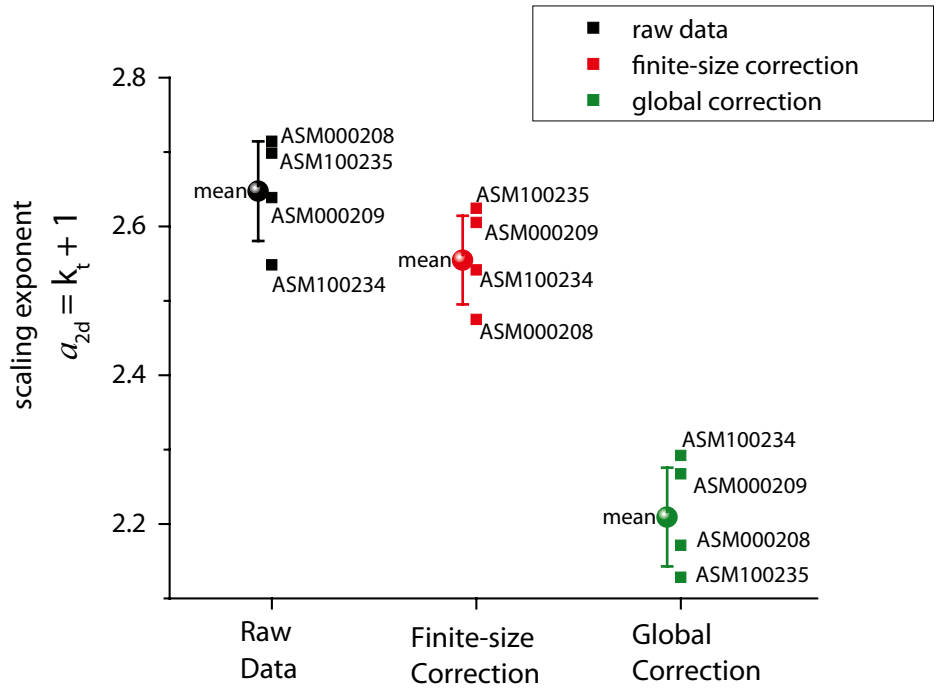


Figure 2-14. *Exponents for Forsmark (a) and Laxemar (b) outcrops before and after correction. Corrected exponents are smaller (around 0.4 units) and less distributed (by approximately two times) than the original exponents. The Global Correction encompasses topography, linkage and finite size effect. Both the linkage and finite size effects contribute to the scaling exponent variation shown.*

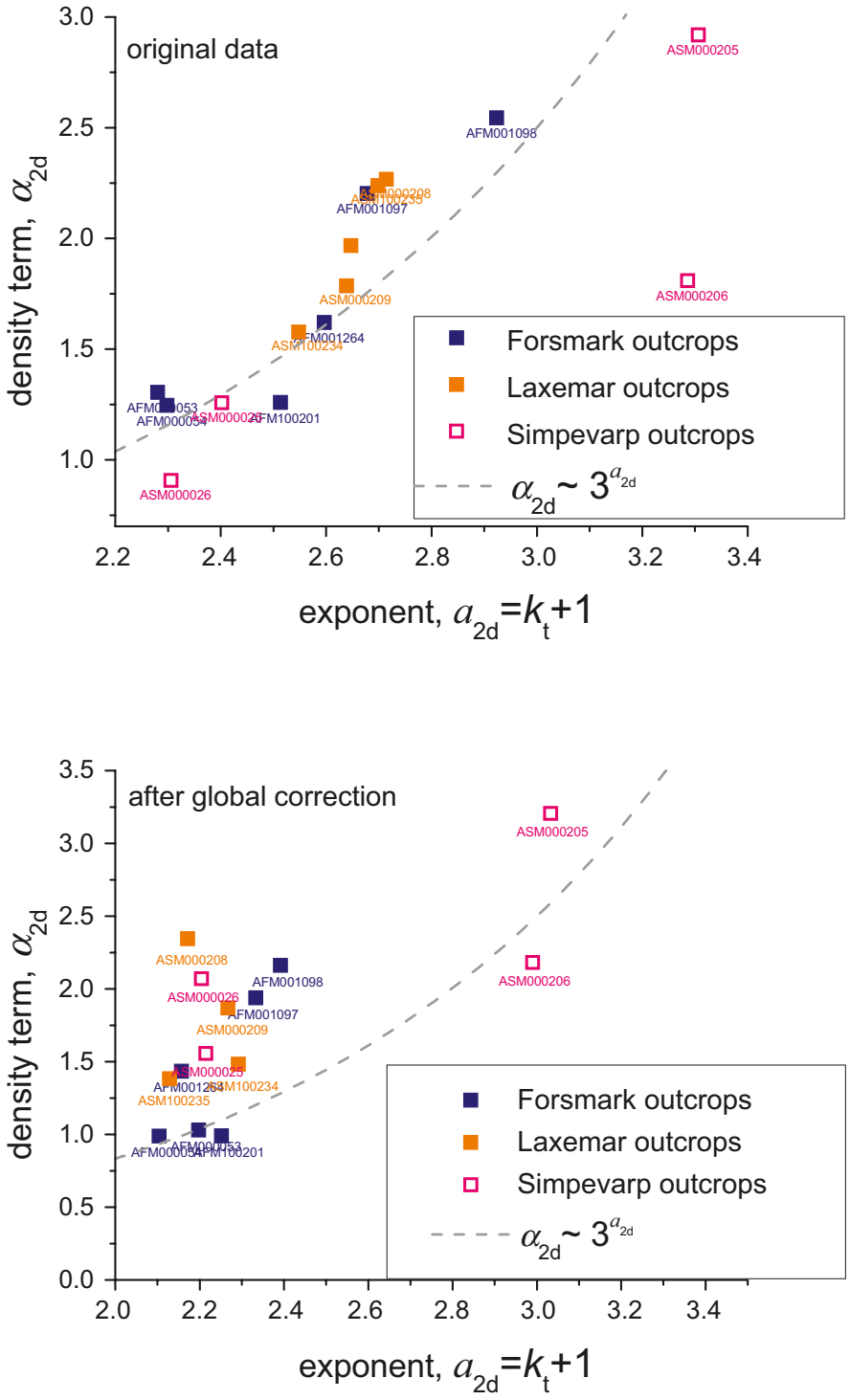


Figure 2-15. *Relation between exponent (k_t) and density term (α_{2d}) before (a) and after (b) data correction.*

This corresponds to the occurrence of one invariant point, at fracture trace size equal to 3, here almost common to all fitted distributions of raw data. After the correction, this trend vanishes. Almost all exponents display values tightened up around k_t equal to 1.2, and the density terms remained quite dispersed. Only two fracture trace maps, ASM000205 and ASM000206 keep a large exponent with k_t very close to 2.

Finally, the mean fracture-size distribution model associated with the respectively Forsmark and Laxemar sites are very close and equal to:

$$n_{Forsmark}(t) = 1.42 \cdot t^{-(1.23+1)}$$

$$n_{Laxemar}(t) = 1.77 \cdot t^{-(1.21+1)}$$

As a general result, on average, outcrop trace-map size distributions are characterized by a power-law model with exponent $k_t = 1.2$. The correction process entailed a significant decrease of the power-law exponent. This eventually increases the gap between the outcrop scale of observation and the lineament scale, where former analyses (/Darcel et al. 2006/ and /Darcel et al. 2004/) ended up with the self-similar model ($k_t = 2$) for these larger scales. We review the lineament size distribution in Section 2.5.3. We will propose a new physical DFN model to encompass the discrepancies between scales in Section 2.6.

2.5.3 Fracture size distribution through scales

All normalized density distributions are combined on multiscale log-log diagrams in Figure 2-16 below. All available scales of observation of DFNs (Figure 2-2) are included, from the outcrop-meter up to the lineament kilometer scales.

Former lineament-map analyses (/Darcel et al. 2006/ and /Darcel et al. 2004/) proposed a scaling exponent, k_t , close to 2, in a model compatible with the "Sweden map scale" (orange in the figures). Note that for the regional scale lineaments at Forsmark (Figure A1-1*d*), it appeared that mapping had been performed both inland and offshore, leading to local variations in resolution scale. In the previous analysis we mentioned that "many trends" could be measured on the density distribution curves, with several possible power-law fits, from values close to –1 (central part of the undersea set) up to –2.95 (largest lineaments, above one km). These trends resulted partly from the mixing of resolution scales. The only trend apparent on both datasets was encountered for the largest scale ($a_{2d} = 2.95$, eq. to $k_t = 1.95$)".

More recent data are available for the present study (the subregional scale datasets displayed in green lines on Figure 2-3*a* and *b*, and Figure A1-1*a* and *b*) : they are made of lineaments defined between 100 m and about 1,000 m. These data are thus intermediate between the trace map and the regional scale lineaments (Figure A1-1*c* and *d*). Their size distributions lead to very different interpretations according to the site under consideration:

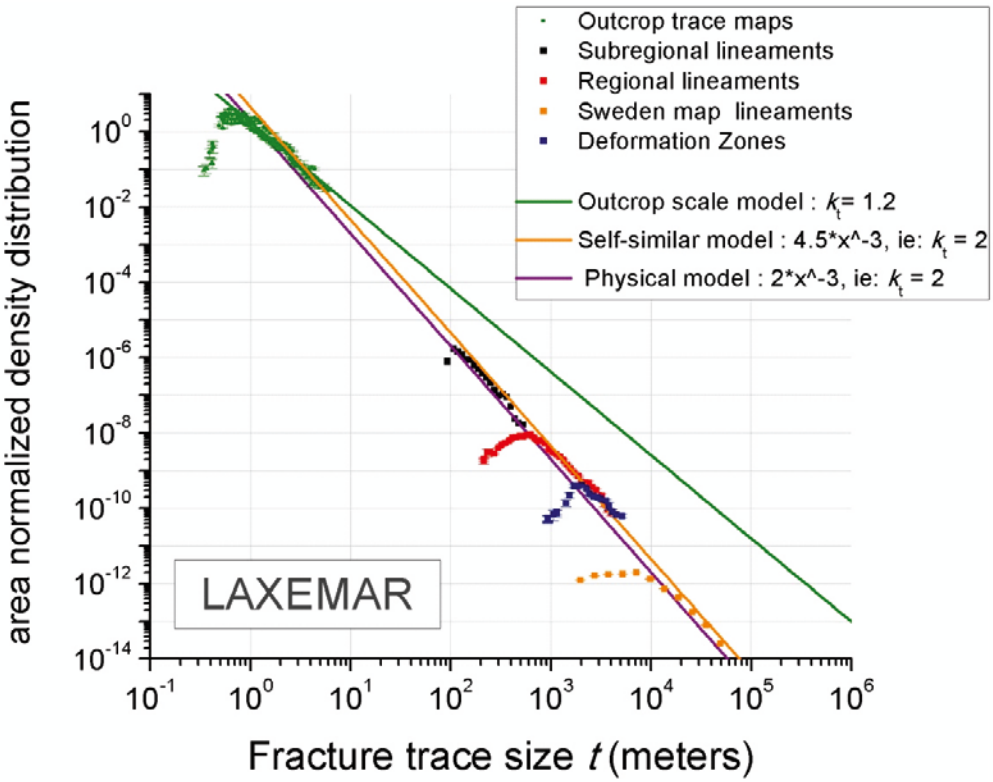
- the Laxemar dataset displays a scaling behavior almost perfectly in agreement with the self-similar scaling ($k_t = 2$).
- the Forsmark dataset displays a dramatic exponent of $k_t = 0.1$. The latter value is highly unusual.

Despite some local discrepancies, the two lineament datasets in Laxemar can be viewed as consistent with the self-similar model originally proposed in /Darcel et al. 2006/, from the hundred meters up to the tens of kilometers scale. Also, the self-similar model intersect the outcrop density distributions at about 3 meters.

On the contrary, the Forsmark lineaments (and DZ) do not follow a simple behavior. As for the Laxemar case, the local outcrop scale model and the self-similar models intersect at about 3–5 meters. Also, the regional scale lineament dataset is consistent with the self-similar model (in its rightmost part). But the sub-regional scale one displays a very different trend. Again, the density distribution intersect the self-similar model at about 400 m, however its local trend radically differs, reflecting a size distribution largely dominated by largest structures. This trend is not observed for the regional scale lineament map. Moreover, a visual inspection of the lineament maps (Figure A1-1*b*) does not help in understanding the variation. It can be proposed that reconnection applied to the sub-regional scale map in Forsmark has been conducted up to an extremely high degree, leading to increase the statistical weight of largest fractures. We note, also, that the Forsmark map contours are extremely irregular, and that resolution seems to vary inside the dataset. We have no definitive conclusion for explaining the density of this dataset.

As a rough conclusion, the scaling behavior of the fracture size distribution follows two major trends: on small scales, an exponent k_t around 1.2; and, on large scales, an exponent of 2.0. In the following section, we propose a physical modeling framework for explaining the two regimes, from k_t smaller than 2 at the meter scale, up to $k_t = 2$ (i.e. the self-similar model) at larger scale.

a)



b)

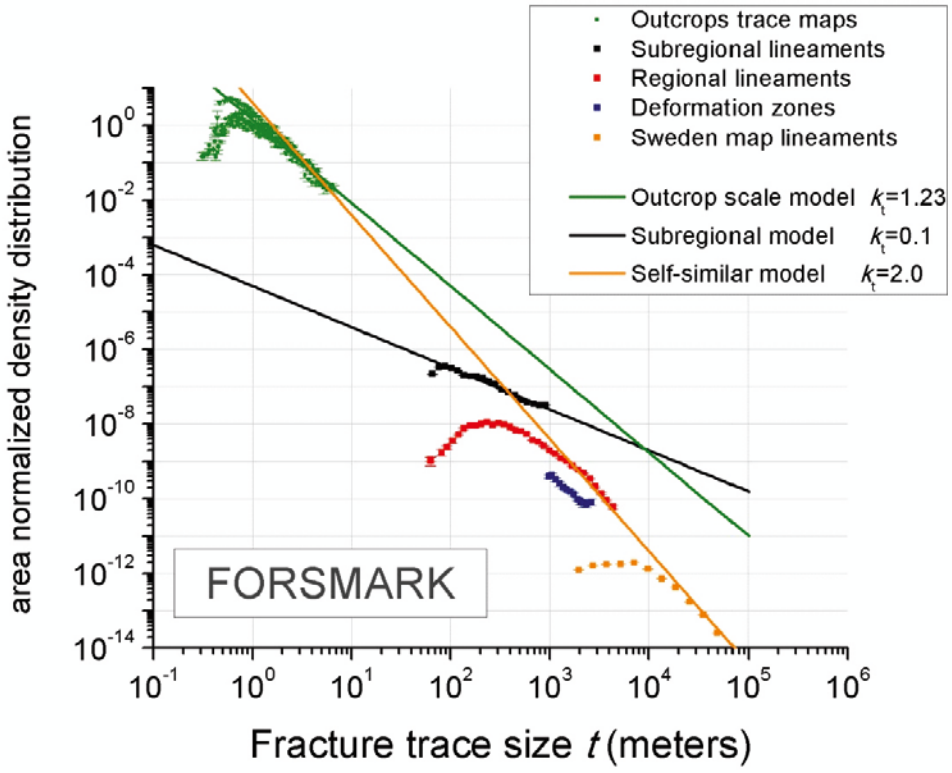


Figure 2-16. *Fracture size distributions for all available datasets, a) Laxemar and b) Forsmark.*

2.6 Physical rationale of a Universal model of Fracture Scaling

The apparent variation in scaling exponent, from the outcrop scale ($k_t = 1.2$) on one side, to the lineament scale ($k_t = 2$) on the other, addresses the issue of the nature of the transition and of its rationale. We develop a new "mechanistic" model that could help in modeling why and where this transition can occur. The model justifies the existence of both distributions and proposes that the "large-scale" power-law derives from general rules that make it a universal fracture scaling model. All the developments presented here are currently submitted to publication in [Davy P, Le Goc R, Darcel C, Bour O, de Dreuzy J R and Munier R, A Universal Model of Fracture Scaling and its consequences for crustal hydro-mechanics].

Basically, our model is based on the fact that there are at least two fracturing regimes. When the density of fracture is low, fractures can freely grow and the eventual length distribution depends on the length dependency of the growth rate. When the fracture density is large, fracture growth is controlled by the mechanical interactions between fractures. The first regime is likely to yield power-law length distribution with "small" exponents k_t about 1 /Sano et al. 1981, Sornette and Davy 1991, Mansfield and Cartright 2001/ although this issue is still debated. We argue that the second regime eventually produces a self-similar power-law length distribution of the form:

$$n(l,L) = \frac{D}{\gamma^D} \cdot L^D \cdot l^{-(D+1)} \qquad 2\text{-}9$$

L is the size of the observation window, D the fractal dimension and γ^D a dimensionless parameter that quantifies both fracture orientations and mechanical interactions. This model was derived from the assumption that a fracture likely stops growing when it encounters a larger one. This puts the emphasis on the inhibiting role played by large fracture on smaller ones. Equ. 2-9 describes the distribution of fractures whose distance to a large one is proportional to fracture length, γ being the proportionality factor. This model called UMFS (for "Universal Model of Fracture Scaling") is self-similar with a size exponent $a_{2d} = k_t + 1 = D_{2d} + 1$ – i.e. $k_t = 2$ if non fractal. The resulting density term, D/γ^D, is dimensionless, and found to lie in the range 2–5 from 2D numerical simulations (or equivalently γ varies from 0.5–1). The UMFS model is consistent with both Laxemar and Forsmark data with $\gamma = 0.7$ and a density term of 4.

The free-growing regime corresponds to small fractures if the corresponding power-law length exponent k_t is smaller than 2. Indeed the average distance between a fracture of length l and any larger fracture increases but at a rate smaller than fracture length. Thus there exists a critical length for which a free growth is no more possible. The transition between both regimes corresponds to a critical fracture length, which depends on the density of fractures in the free-growing regime. This is illustrated Figure 2-17, which depicts a simulation based on an initial model with $\alpha_{2d} = 0.53$, $k_t = 1.3$, two families with perpendicular orientations, and where fractures are truncated according to larger ones. The initial underlying model remains valid *a posteriori* up to a critical scale ($l_c = 7$). Above this critical scale, the Universal Self-Similar Scaling model is reached with $k_t = 2$ and $\alpha_{2d} = 2$. Parameters of the UMFS are independent of the initial size distribution: varying the initial a_{2d} and α_{2d} will modify only the critical scale from which the universal trend is reached.

How this theory applies to the Forsmark and Laxemar sites? At the outcrop scale, the mean fracture-size distribution model for both these sites is about similar with an exponent k of about 1.2 and a density term that varies between 1.0 and 2.5. At the lineament scale (> 100 m) and for a few outcrops, the self-similar model is likely valid. These observations are consistent with the aforementioned theory. With these parameters, the transition between both regimes would occur for a fracture length of 1–10 m and even at smaller scale for the few outcrops that follow the self-similar density model. A consequence for the disposal issue is that the model that is likely to apply in the "blind" scale window between 10–100 m is the self-similar model as it is defined for large-scale lineaments.

This above discussion does not bring factual evidences for a general distribution model. It just provides a rationale for linking small-scale to large-scale distributions. The self-similar model, as it is measured for some outcrops and most lineament maps, is definitely worth being investigated as a reference for scales above 1–10 m.

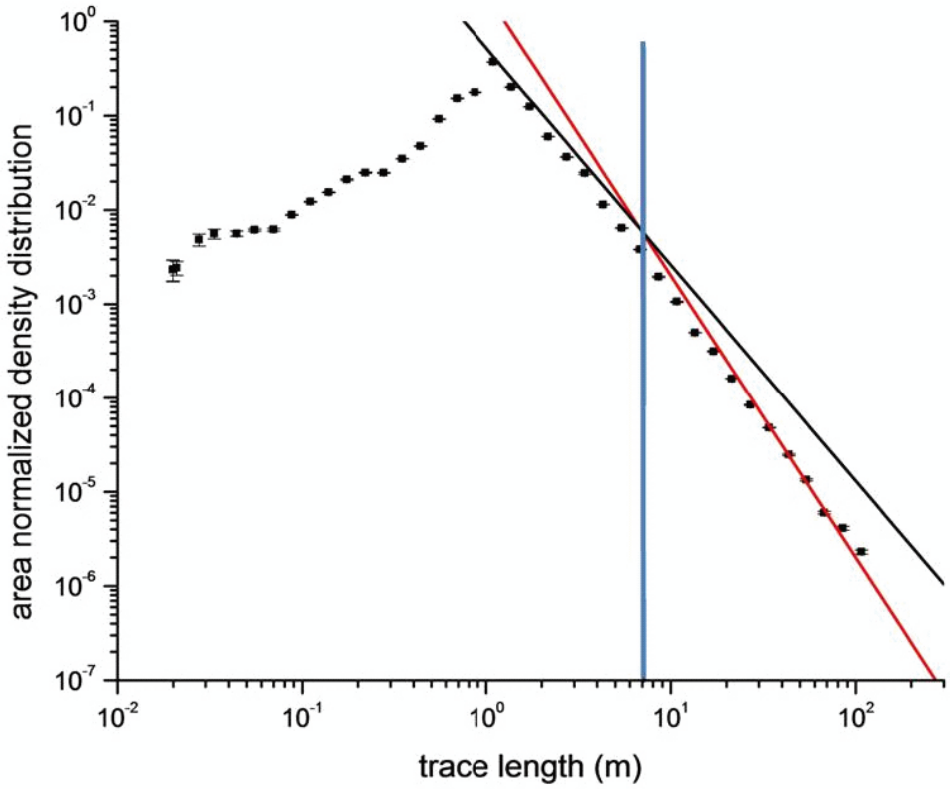


Figure 2-17. *Numerical illustration of the UMFS with simulated fracture-size distribution. The simulation arising from the numerical model is represented with black symbols, the underlying input generation model as a black straight line, and the resulting self-similar model at larger scale as a red straight line. We observe here the match between the simulation and the expected model with a change of behavior at* $\mathbf{l}_c = 7$.

3 Data resources at depth and dataset definitions

3.1 Data delivery

In addition to the delivery of surface data (section 2.2 and Appendix 1 Data resources for the fracture size analyses), 3 files from SICADA are used in the depth analysis (See Appendix 2):

– P_frac_core_eshi
– Object location
– P_freq_1m_eshi

3.2 Elementary definition of datasets

When conducting DFN modeling, we first need to ascertain which geological defects are fractures and what makes a fracture. This aspect of the problem is handled, for the surface data, through the linkage of apparently segmented fracture traces (section 2.4.2). At depth, for borehole data, the DFN basis is made of the fracture intercepts identified and stored in the **p_frac_core_eshi** database.

We could have considered fractures from the sealed network and crush zones to this database. However, in the quantitative analyses presented hereafter, sealed networks and crush zones are not yet included. Their quantitative combination with the other set of individually defined fractures is not obvious. Data associated to a given crush zone or sealed network zone are provided as *i*) start and end position of the zone, *ii*) fracture density in the zone and sometimes *iii*) a global corresponding orientation. To use these fractures would mean to add them directly in the form of clusters or groups, with orientations poorly defined. This would add strong uncertainties to the final density/orientation DFN model. An alternative possibility would be to consider separately a macro Discrete "Large Structures" Network model, including only sealed networks and/or crush zones, or even Deformation Zones (Section 7.3.1 in /Darcel et al. 2006/).

As recommended by SKB, we consider DFN modeling within Fracture Domains, as opposed to Deformation Zones Domains (DZ). As illustrated in Figure 3-1 at borehole KFM01C (from Forsmark), the investigated site is basically divided, into DZs (the top-most layer in the figure), where DFN models are not expected, and remaining areas, in which relevant DFN models are expected. The latter zones also are subdivided into various subdomains according to rock, geology or fracturing properties. The crush zones and sealed network zones overlap with the preceding domain identifications. Following the SKB recommendations, all the fracture intercepts located in the DZs are "erased from the analysis"; when this occurs, DZ portions along boreholes are removed, and remaining pieces are stacked together. Sticking the sub-pieces together is essential to the scale-variability analysis.

At the Forsmark site, a total of 10 main FDs (FFM01 to FFM06, plus additional "affected by DZ", /Olofsson et al. 2007/) are defined. The repartition in the FD at the borehole scale can be quite complex, as illustrated in Figure 3-1*b* (↔FD), with a succession, from top to bottom, of FFM02, FFM02_ affected_by_DZ, again FFM02, and then FFM01_affected_by_DZ. This organization underlines the difficulty to share a site into consistent domains, especially in the context of multi-scale fracturing. At the Laxemar site, a total of 6 FDs are defined, named FSM_W, FSM_C, FSM_EW0007, FSM_N, FSM_NE005 and FSM_S.

As already mentioned in Introduction, we perform in the following an analysis at the site scale, including simultaneously several FDs, rather than FD by FD. As a consequence, the elementary datasets are defined borehole by borehole and according to the local Fracture Domain divisions (as an example, "KFM01C_ FFM01_affected_by_DZ" is one dataset). Moreover, if one Fracture Domain occurs in discontinuous parts along a borehole (after DZ have been removed), the dataset is further divided (this happens rarely and only at Forsmark). Therefore, most of the times, one dataset is fully defined by reference to its borehole and Fracture Domain. By doing this, we come to define 60 datasets at Forsmark and 51 at Laxemar. The number of records (fracture number) and the section size to which they are related are provided for each dataset in Figure 3-2. A total of 16.5 km at Forsmark, and 17.5 km at Laxemar of cored borehole are available for analysis, including, respectively, close to 90,000 and 106,000 fracture records. Because they are built from the FD division, we expect, *a priori*, a relative consistency between various datasets from a single FD.

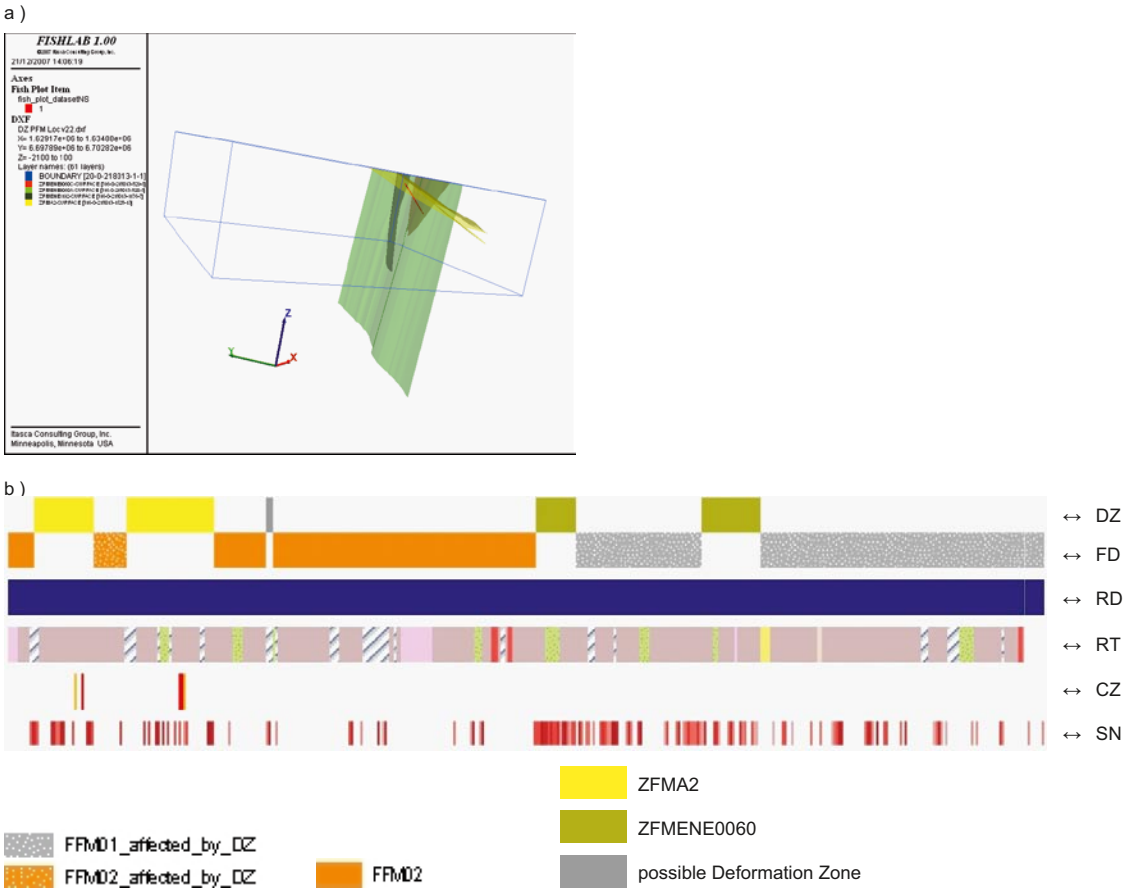


Figure 3-1. ***a)*** *3D representation of borehole KFM01C (thick red line) with modeled DZ intersecting it (from DZ PFM Loc V22).* ***b)*** *Schematic representation of borehole KFM01C geological interpretations – From top layer to bottom layer: Deformation Zones, Fracture Domains, Rock Domains, Rock Types, Crush Zones and Sealed Network zones); depth increases toward right (→).*

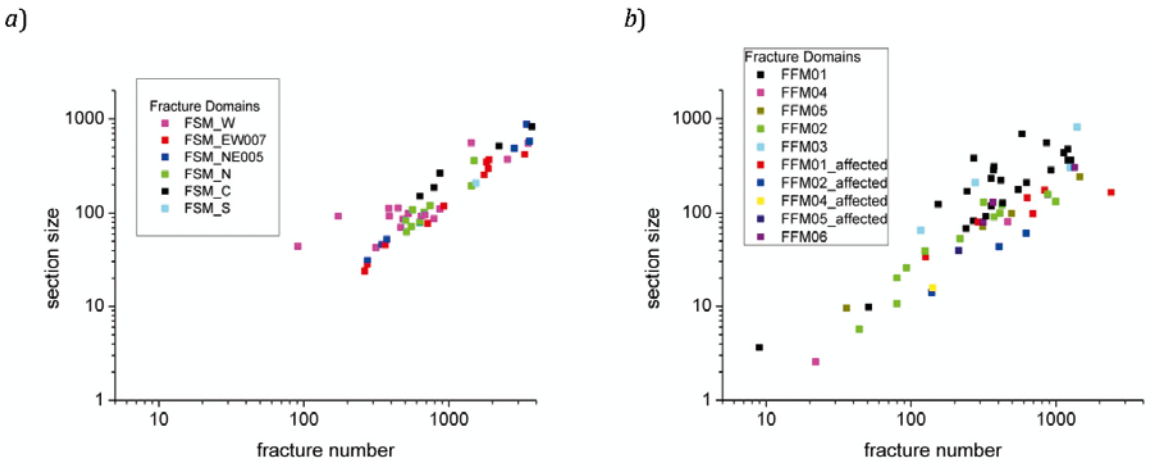


Figure 3-2. *General dataset characteristics in terms of number of fracture intercepts recorded (x-axis) and borehole section size (y-axis) for* ***a)*** *Laxemar and* ***b)*** *Forsmark.*

4 Including variability and uncertainty in modeling

4.1 Introduction

In this section we develop quantitative tools to characterize the variability and uncertainty associated to DFNs (issues 4 and 6 in the Introduction). By doing this, it becomes possible to assess whether available data are sufficient or not to realistically characterize fracturing properties and define a site DFN model.

In the present work, we focus on uncertainties relative to fracture pole orientations. Therefore, the development described here is applied to typical borehole data, and focused on the density and orientation properties of DFN models. In order to analyze their impacts on DFN characterization, the uncertainty, σ_{unc} , and the variability, σ_{var}, associated to density distributions, are integrated into the modeling framework, to be propagated further into a remaining/final uncertainty associated to the DFN model parameters.

These developments are the necessary preamble for the statistical FD analysis entailed in Section 5. The chapter is organized as follows: (a)a recall of the definition and model framework; (b) the propagation of orientation uncertainties up to the density/orientation distribution model; (c) the variability assessment; and (d) synthesis to compare the relative weight of σ_{unc} and σ_{var} (applied over both sites).

4.2 Definitions and model framework

4.2.1 1D DFN distribution to parent distribution

The statistical DFN model, $N_{3d}(L, l, \theta, \varphi)$, provides the number of fractures in a given volume with given sizes (l) and orientations (strike θ and dip φ) [notations similar to /Darcel et al. 2006/]. As explained in Section 4.2 of /Darcel et al. 2006/, using the scaling assumptions leads to:

$$N_{3d}(L,l,\theta,\varphi) = \alpha_{3d}(\theta,\varphi)\cdot l^{-(k_r+1)}\cdot L^{D_{3d}} \qquad 4\text{-}1$$

The **normalized density distribution model** is simply the above resulting number divided by the volume:

$$n_{3d}(l,\theta,\varphi) = \frac{N_{3d}(L,l,\theta,\varphi)}{L^{D_{3d}}} \qquad 4\text{-}2$$

The density term $\alpha_{3d}(\theta, \varphi)$ represents the fracture density for a fracture size of 1.

Along a borehole, a dataset made of fracture intercept records is assimilated to a one-dimensional (1D) distribution of fractures, noted as $n_{1d}(\theta, \varphi)$. It simply provides the number of fracture intercepts per unit of borehole length; the fracture intercept pole orientation is (θ, φ). Although the fracture sizes cannot be measured directly from the borehole data, their associated size distribution is part of the underlying model. A stereological relation between parent and observed models can be derived. In case of a power-law radius distribution of fractures in 3D, $n_{1d}(\theta, \varphi)$ and $\alpha_{3d}(\theta,\varphi)$ are related through the following equation /Davy et al. 2006/:

$$n_{1d}(\theta,\varphi) = \alpha_{3d}(\theta,\varphi)\cdot f(r_{min},k_r,h,d)\cdot P(k_r,\hat{a}) \qquad 4\text{-}3$$

where r_{min} and k_r are the DFN model parameters, h is the borehole length, d is the borehole diameter, $\hat{a}$ is the acute angle (the angle between the fracture pole and borehole direction) and (θ,φ) the fracture pole orientation. $P(k_r,\hat{a})$ represents a correction term related to orientation bias (Eqs. 6 and 7 in /Davy et al. 2006/), and $f(r_{min},k_r,h,d)$, a factor combining the intersection between the 3D DFN and the borehole. We recall that for $k_r > 3$ and $r_{min} < d/2$; then, $P(k_r,\hat{a}) \sim \sin(\hat{a})^{kr-1}$, where $\hat{a}$ is the acute angle between the fracture pole and the borehole; in other cases, it is such that $P(k_r,\hat{a}) \sim \sin(\hat{a})$. In order to simplify equation-reading, the notation below will be used:

$$\psi(\hat{a}) \equiv f(r_{min},k_r,h,d)\cdot P(k_r,\hat{a}) \qquad 4\text{-}4$$

Occurrence and pole orientation of fractures are necessary to fully compute the initial apparent density distribution $n_{1d}(\theta, \varphi)$ and, then the parent density distribution, $\alpha_{3d}(\theta,\varphi)$. If orientation information is missing or partial some assumptions must be done in order to define the parent density. The case of sealed networks and crush zones is explained in section 3.2 above. They have not been included in analyses. Despite orientation uncertainties are higher for fracture Not Visible In BIPS, their occurrence and orientations are defined. They are therefore explicitly included in the analyses. Note that we have also checked an approximated model where only the occurrence of all fractures is computed but with orientations arising from only the Not Visible In BIPS. By comparing the full and approximate computation, we have concluded that the relatively high uncertainties associated to Not Visible In BIPS can be exactly accounted for.

Finally, density terms α_{3d} can be replaced alternatively by the more usual density parameter P_{32} (adapted from equations on p. 15 in /Darcel et al. 2006/):

$$P_{32}(\theta,\varphi;l_c) = \frac{1}{L^3}\frac{\pi}{4}\int_{l_c}^{L}\left(\alpha_{3d}(\theta,\varphi)\cdot l^{-(k_r+1)}\cdot L^D\right)\cdot\pi\cdot\left(\frac{l}{2}\right)^2 dl \cong \frac{\pi}{4}\cdot\alpha_{3d}(\theta,\varphi)\cdot L^{D-3}\cdot\frac{\left[l_c^{-k_r+2}\right]}{k_r-2} \qquad 4\text{-}5$$

4.2.2 Density/orientation computations

The density terms mentioned above are functions of orientation: in practice, the (θ,φ) in $\alpha_{3d}(\theta,\varphi)$ refers to the subdivision of a dataset according to orientation.

The binning over orientations reflects the degree of detail given to the relation between fracture density and orientations. It can be used to discriminate between fracture sets. However, its level of detail relies greatly on the necessity to have a statistically relevant number of fracture in the bins, this to be able to derive a sound calculation of $\alpha_{3d}(\theta,\varphi)$. In practice we consider three grids with varying level of details. A bin of a grid can be called a angular or pole zone. The first case is not really a grid, as it encompasses all orientations in a single dataset (Figure 4-1*a*). In this case, we obtain α_{3d}, such:

$$\alpha_{3d} = \iint_{\theta,\varphi}\alpha_{3d}(\theta,\varphi)\cdot d\theta\cdot d\varphi \qquad 4\text{-}6$$

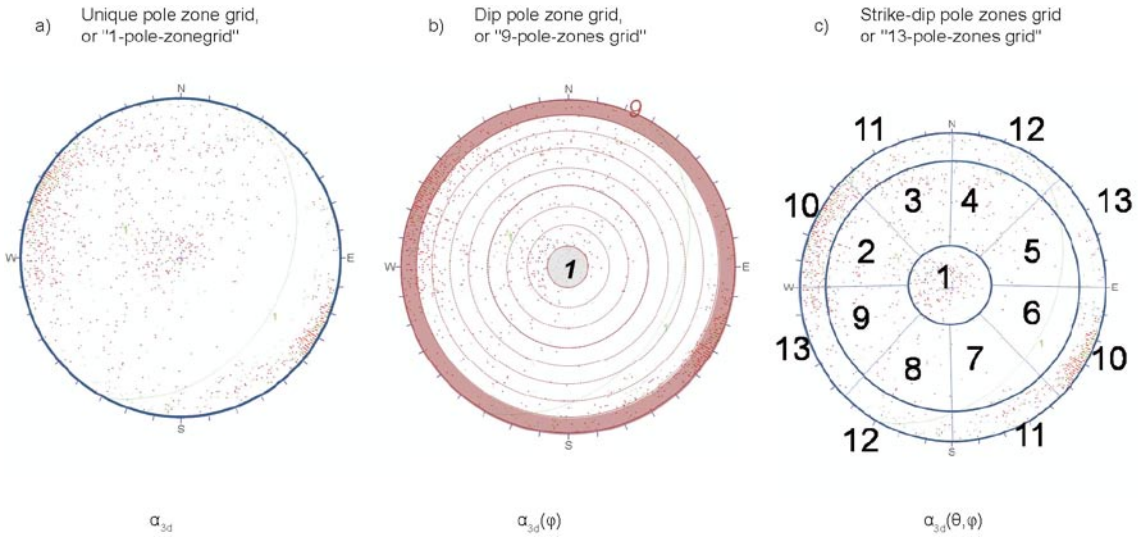


Figure 4-1. *Illustration of three main grids of analysis in a stereographic representation (Red points refer to fracture poles.) Case a: No analysis with respect to orientation is done; all fracture poles are considered together. Case b: Bins are defined according to dip, spacing is 10°, zone 1 and 9 are highlighted in the figure. Case c: A complete binning over 13 pole zones. (Discretization of the (θ,φ) space in 13 pole zones. Dip is taken in the intervals [0°, 22.62°] , [22.62°, 72.08°] and [72.08°, 90°], whereas strike is taken in 45°-width intervals between 0° and 360°)*

The second represents densities according to dip ranges:

$$\alpha_{3d}(\varphi) = \int_{\theta} \alpha_{3d}(\theta,\varphi)\cdot d\theta \qquad 4\text{-}7$$

Finally, any other subdivision in the "pole space" can be considered. The most detailed selected division encompasses 13 angular zones, distributed as illustrated in Figure 4-1*c*. Each of these zones are denoted as a "pole zone" (in order to avoid confusion with other zones, domains, etc).

The three cases mentioned above lead to the observation of the fracturing properties through various points of view: Case a, considering only the density; Case b, considering only the relative proportion of vertical versus horizontal fractures through dip dependence; or Case c, considering regular bins.

4.3 Density and Uncertainty modeling

4.3.1 Introduction

Here, **Uncertainty** arises from **measure precision**. We show how it is propagated into the DFN density distribution model. Accordingly, the density term $\alpha_{3d}(\theta,\varphi)$ is associated with an uncertainty, noted $\sigma_{unc}(\theta,\varphi)$, such that the effective density associated to a dataset is simply $\alpha_{3d}(\theta,\varphi) \pm \sigma_{unc}(\theta,\varphi)$. $\alpha_{3d}(\theta,\varphi)$ and $\sigma_{unc}(\theta,\varphi)$ values are computed through a two steps procedure. It encompasses:

– the transformation (see section 4.3.3) of uncertainties from local angles α and β, initially defined relatively to borehole direction /Munier and Stigsson 2007/, into uncertainties associated to global angles, the classical strike and dip (θ,φ).

– the main step that consists in going from the discrete data (fracture intercept parameters, including uncertainties on pole orientations) up to the parent density distributions. See section 4.3.2.

This two-step procedure and its numerical validation (section 4.3.4) are developed in the following sections.

4.3.2 A 3D density model from discrete datasets

We note the following:

- A normalized, centered, Gaussian law can be associated with each fracture measure (φ_i, θ_i, $\hat{a}_i$, $\sigma_{\varphi,i}$, $\sigma_{\theta,i}$, $\sigma_{\hat{a},i}$), with φ_i, θ_i and $\hat{a}_i$ representing the means, and $\sigma_{\varphi,i}$, $\sigma_{\theta,i}$ and $\sigma_{\hat{a},i}$ representing the standard deviations. Subscript *i* simply refers to fracture *i* in the dataset.
- A probabilistic approach allows $\alpha_{3d}(\theta,\varphi) \pm \sigma_{unc}(\theta,\varphi)$ to be defined from discrete data.

We consider the probability for a discrete record to have its dip value within a given range ($\varphi_1;\varphi_2$), as schematized in Figure 4-2. There are only two possibilities: either pole φ is within or outside the range, which leads to consideration of event *C*:

$$C = \begin{matrix} 1 \text{ if } \varphi \in [\varphi_1,\varphi_2] \\ 0 \text{ if } \varphi \notin [\varphi_1,\varphi_2] \end{matrix}$$

The mean of $< C >$ and its standard deviation, σ_C , are then

$$\begin{aligned} \langle C \rangle &= p_{\varphi\in[\varphi_1,\varphi_2]} \\ \sigma_C &= \sqrt{p(1-p)} \end{aligned} \qquad 4\text{-}8$$

where *p* is the probability for a fracture *i* of belonging to a given interval. Considering the Gaussian distribution associated with orientation of fracture *i*, and φ_1 and φ_2, the interval boundaries, one gets

$$p\left(\varphi_i \in [\varphi_1,\varphi_2]\right) = \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma_{\varphi_i}}\int_{\varphi_1}^{\varphi_2} e^{-\left(\frac{\varphi-\varphi_i}{\sqrt{2}\cdot\sigma_{\varphi_i}}\right)^2}\cdot d\varphi \qquad 4\text{-}9$$

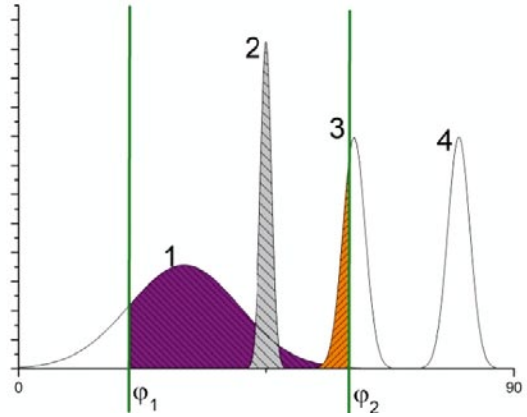


Figure 4-2. *Illustration on the contribution of a fracture intercept to a delimited range [φ_1 ; φ_2]. Here cases 1 to 3 contribute to the range [φ_1 ; φ_2] whereas case 4 has a negligible contribution on the density over that range. The contribution of each intercept is computed through integration (dashed colored zones) of its associated Gaussian distribution (Equ. 4-9). The mean and standard deviation are dependent on the fracture data.*

When both angles defining the fracture pole orientation are considered,

$$\left\langle C_{\theta,\varphi} \right\rangle = p_\theta \cdot p_\varphi = p_{\theta,\varphi}$$
$$\sigma_C^2 = p_{\theta,\varphi} \cdot \left(1 - p_{\theta,\varphi}\right)$$

The probability p_i , of fracture *i* , belonging to the complete orientation range { θ_1, θ_2 ; φ_1, φ_2}, is given by

$$p_i = p\left(\varphi_i \in [\varphi_1, \varphi_2] \cap \theta_i \in [\theta_1, \theta_2]\right) = \frac{erf\left(\frac{\varphi_2 - \varphi_i}{\sigma_{\varphi,i}\sqrt{2}}\right) - erf\left(\frac{\varphi_1 - \varphi_i}{\sigma_{\varphi,i}\sqrt{2}}\right)}{2} \cdot \frac{erf\left(\frac{\theta_2 - \theta_i}{\sigma_{\theta,i}\sqrt{2}}\right) - erf\left(\frac{\theta_1 - \theta_i}{\sigma_{\theta,i}\sqrt{2}}\right)}{2} \qquad 4\text{-}10$$

Finally, the full process encompasses strike and dip angles, as well as the orientation bias correction. The event *J* refers to the contribution of one fracture to the parent density distribution. Then, with *F* being the orientation bias correction (Equ. 4-4) and assuming a weak correlation between *C* and *F*, *J* is written as

$$\left\langle J \right\rangle = \left\langle C \right\rangle \cdot \left\langle F \right\rangle$$
$$\sigma_J^2 = F^2\sigma_C^2 + C^2\sigma_F^2 \qquad 4\text{-}11$$
$$\text{"with } \sigma_F^2 = \left(\frac{d\,\mathrm{F}}{d\hat{a}}\sigma_{\hat{a}}\right)^2 \text{ and } F \equiv \psi(\hat{a})\text{"}$$

Accordingly, for a dataset containing N records, the contribution of each fracture intercept *i*, to each term of the density distribution, is computed thanks to the probabilities above defined and leads to:

$$\alpha_{3d}\left([\varphi_1, \varphi_2], [\theta_1, \theta_2]\right) = \sum_{i=1}^{N}\left[p_i \cdot \psi\left(\hat{a}\right)\right]$$
$$\sigma_{unc}{}^2\left([\varphi_1, \varphi_2], [\theta_1, \theta_2]\right) = \sum_{1}^{N}\left[p_i(1-p_i)\cdot\psi\left(\hat{a}_i\right)^2 + \left(p_i \cdot \frac{\mathrm{d}\psi}{\mathrm{d}\hat{a}} \cdot (\hat{a}_i)\sigma_{\hat{a}_i}\right)^2\right] \qquad 4\text{-}12$$
$$\text{with } p_i = p\left(\left(\varphi_i \in [\varphi_1, \varphi_2]\right) \cap \left(\theta_i \in [\theta_1, \theta_2]\right)\right)$$

Uncertainties related to the fracture intercept orientation thus contribute to the orientation/density distribution and to the bias correction intensity. If no binning is defined (only one class of orientation where all fractures contribute), the orientation intensity reflects only on the density term.

In summary, the 3D density/orientation distribution model, $n_{3d}(\theta,\varphi)$, is computed:

a) by assuming a normalized, centered Gaussian law for each fracture intercept measure (φ_i, θ_i, $\hat{a}_i$, $\sigma_{\varphi i}$, $\sigma_{\theta i}$, $\sigma_{\hat{a}i}$);

b) from an initial set of discrete fracture intercepts distributed over a borehole section of size h, noted $\{(\varphi_i, \theta_i, \hat{a}_i, \sigma_{\varphi i}, \sigma_{\theta i}, \sigma_{\hat{a}i}; h)\}_{i=1...N}$. Note that, once a fracture pole orientation is expressed through strike and dip values, the acute angle $\hat{a}$ and its uncertainty $\sigma_{\hat{a}}$ still are used to calculate the orientation bias correction and the related uncertainty.

c) for any range of strike/dip values. This is equivalent to calculating different densities for different fracture sets once boundaries are fixed in the "orientation space".

4.3.3 Strike and dip uncertainty from fracture poles

As written in /Munier and Stigsson 2007/, the strike (θ) and dip (φ) of fractures and other rock structures are calculated using the local alpha (α) and beta (β) angles together with the bearing (B) and inclination (I) of the borehole at actual depth.

The fracture pole orientation conversion from ($\hat{a}$, β; B, I) angles to strike and dip (θ, φ) is described in Eqs. 4 through 8 in /Munier and Stigsson 2007/. These equations are derived in order to obtain the uncertainties on strike (σ_θ) and dip (σ_φ):

$$\langle f \rangle = f(\hat{a}_i, \beta_i, I_i, B_i)$$

$$\sigma_f{}^2 = \frac{1}{2}\cdot\left[\left(\left.\frac{\partial f(\hat{a},\beta,I,B)}{\partial \hat{a}}\right|_{\hat{a}_i,\beta_i,I_i,B_i}\cdot\sigma_\alpha\right)^2+\left(\left.\frac{\partial f(\hat{a},\beta,I,B)}{\partial \beta}\right|_{\hat{a}_i,\beta_i,I_i,B_i}\cdot\sigma_\beta\right)^2\right.$$
$$\left.+\left(\left.\frac{\partial f(\hat{a},\beta,I,B)}{\partial I}\right|_{\hat{a}_i,\beta_i,I_i,B_i}\cdot\sigma_I\right)^2+\left(\left.\frac{\partial f(\hat{a},\beta,I,B)}{\partial B}\right|_{\hat{a}_i,\beta_i,I_i,B_i}\cdot\sigma_B\right)^2\right] \qquad 4\text{-}13$$

where f represents either the strike θ or the dip φ (resp Eq. 5 or Eq. 6 in /Munier and Stigsson 2007/) of a fracture intercept denoted by subscript "$_i$".

4.3.4 Validation

Theoretical developments, leading to Equ. 4-13, and Equ. 4-4, describing the density model through its mean value and associated standard deviation, are compared directly to numerical simulations. Essentially, the standard deviation, $\sigma_{unc}(\theta,\varphi)$, associated with $\alpha_{3d}(\theta,\varphi)$ is recomputed through direct simulations using a Monte Carlo approach.

$\sigma_{unc,simulated}(\theta,\varphi)$ is directly computed as the standard deviation of a large number of stochastic realizations of $\alpha_{3d}(\theta,\varphi)$. Each simulated density distribution is generated by random choice in the original dataset, which is based on local angles: $\{\hat{a}_i, \beta_i, B_i$ and $I_i\}_{i=1...N}$. For each fracture pole i, expressed in ($\hat{a}_i$, β_i; I_i, B_i) local angles, a pole orientation is randomly picked up in the associated set of Gaussian distributions (i.e. for the $\hat{a}$ angle, the mean is $\hat{a}_i$, the standard deviation is $\sigma_{\hat{a}i}$, etc). **This is equivalent to exactly sample the dataset over a range ($\hat{a}_i \pm \sigma_{\hat{a}i}/2$; $\beta_i \pm \sigma_{\beta i}/2$) for each fracture pole i**. The $\sigma_{unc,computed}(\theta,\varphi)$ and $\sigma_{unc,simulated}(\theta,\varphi)$finally can be compared (Figure 4-3).

Data from FD FFM01 at borehole KFM01C are chosen to illustrate the verification process. The total considered borehole section starts at a depth of 170 m and ends at a depth of 371 m. It is divided into several subpieces of various sizes, leading thus to the creation of data subsets of varying weight (total number of data in the set) and varying densities (total number of data in the set divided by the subpiece size). The subsets thus built are listed in Table 4-1. In addition, the chosen pole zone grid is equivalent to defining 13 pole zones according to strike and dip ranges (Figure 4-1*c*).

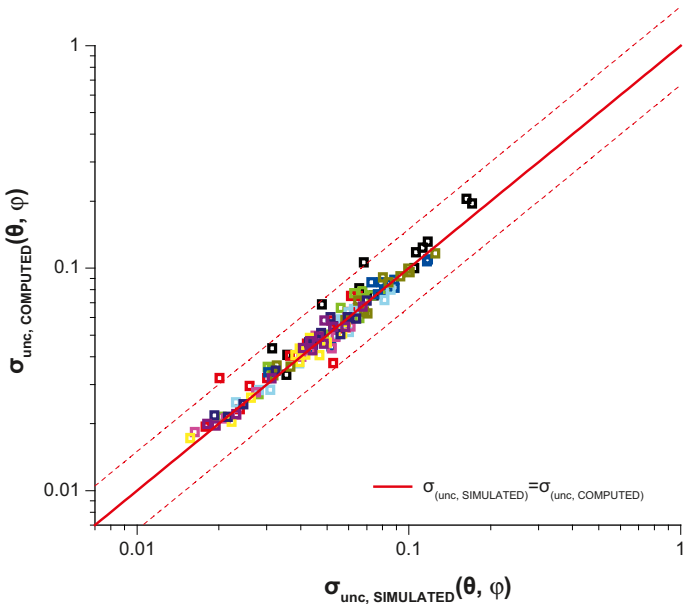


Figure 4-3. *Comparison between $\sigma_{unc,computed}(\theta,\varphi)$ from Equ. 4-12 and $\sigma_{unc,simulated}(\theta,\varphi)$ generated through Monte Carlo simulations. Symbol colors correspond to the various subsections defined in Table 4-1; each subsection is composed of 13 points, reflecting the 13 pole zones of the grid (Figure 4-1c).*

Table 4-1 : Fracture subsets defined according to depth position from borehole KFM01C, FD FFM01.

Set index	Depth min (m)	Depth max (m)	Section size (m)	Number of fractures in the section
1	170	221	51	609
2	170	271	101	1,258
3	170	321	151	1,662
4	170	344	177	1,746
6	220	271	51	660
7	220	321	101	1,064
8	220	344	124	1,148
10	270	321	51	421
11	270	344	74	733
13	320	344	24	89

The Monte Carlo process is based on 100 simulations associated with each dataset. For each subset, the values $\sigma_{unc,computed}(\theta,\varphi)$ and $\sigma_{unc,simulated}(\theta,\varphi)$ are computed; these are presented in Figure 4-3: values from the direct simulations are plotted along the x-axis, and values from the quantitative/theoretical method along the y-axis. The cluster of points thus generated is located along the identity function, reflecting the compatibility between $\sigma_{unc,computed}(\theta,\varphi)$ and $\sigma_{unc,simulated}(\theta,\varphi)$. This validates the method developed to derive $\sigma_{unc}(\theta, \varphi)$ quantitatively.

4.3.5 Maximum bias correction

According to Equ. 4-4 and Equ. 4-12, the orientation bias correction depends mainly on $1/\sin(\hat{a})^{kr-1}$ or $1/\sin(\hat{a})$, for the density term α_{3d} and standard deviation (σ_{unc}) computations. Also, large values of σ_{unc} arise from large values of uncertainties combined to strong orientation bias correction.

Very small acute angles lead to very large corrections of orientation bias. The limit is reached when $\hat{a}$ equals to 0, which leads to an infinite correction. Classically, the greatest applied correction is limited to $\hat{a}$ close to 10–15° in order to avoid "artificially high corrections" due to difficulty in getting a statistical relevancy based on very few data. However, in theory, the probability to catch fracture poles with very low acute angles is low, and must be associated with a bias correction of strong orientation.

In the present development, the maximum allowed correction is increased to very large values such that large corrections occur. We emphasize that the resulting uncertainty, due to statistical weakness (very few poles strongly corrected) also is quantified fully, sometimes leading to values of $\boldsymbol{\sigma_{unc}}$ close to, or even larger than, density α_{3d} itself. In other words, when such large corrections occur on the density term, they also are reflected in the increase of uncertainty.

The effect of the limit value for the orientation bias correction is illustrated in Figure 4-4. The example chosen from the KLX09C_FSM_N dataset is one of the least favorable at Laxemar. The threshold observed in Figure 4-4*a* illustrates the change in conditions below and above a correction equal to 300: below 300, the threshold acts as an arbitrary limit; above 300, it becomes constant, as all the fracture pole values are encompassed into the process. The pending consequence is that uncertainties (standard deviations) become so high that final mean densities are meaningless. In the present example, KLX09C_FSM_N, the density increases from 11.5 to 15.5, but the largest densities (the green symbols in Figure 4-4*b*) lead to uncertainties larger than the value itself. In other words, the difficulty in inferring useful parent distribution parameters from data associated with strong orientation bias still is not resolved, but we give it a clear theoretical framework: data are no longer underestimated, and their degrees of confidence are calculated.

In summary, the process presented here gives rise to larger density values and associated standard deviations. Finally, whenever $\boldsymbol{\sigma_{unc}}$ is computed to be larger or equal to density, the value can be simplified:

$$\sigma_{unc} = \max(\alpha_{3d}; \sigma_{unc,computed})$$

4.4 Application Example

Results from the parent distribution computation [$n_{3d}(\theta,\varphi)$; $\sigma_{unc}(\theta,\varphi)$] are displayed in Figure 4-5. Recall that the α_{3d} density parameter expression can be replaced by the more usual density term, P_{32}, which reflects the total surface of fracture by unit of volume (independent of the size distribution shape, Equ. 4-5). We chose this simplified representation for diagrams in the rest of the report.

In this example (Figure 4-5), a parent distribution is computed on a *13-pole-zones grid* for the orientation subdivision. (Other grids are illustrated in Figure 4-1). As illustrated in the related stereonet representation, the P_{32} associated with the fracture pole belonging to pole zone 12 is close to 1.16, with a standard deviation equal to 0.23. To enable direct comparison of the densities between the various pole zones, all densities and standard deviations are displayed on an x-y diagram with the pole zone reference number in the x-axis.

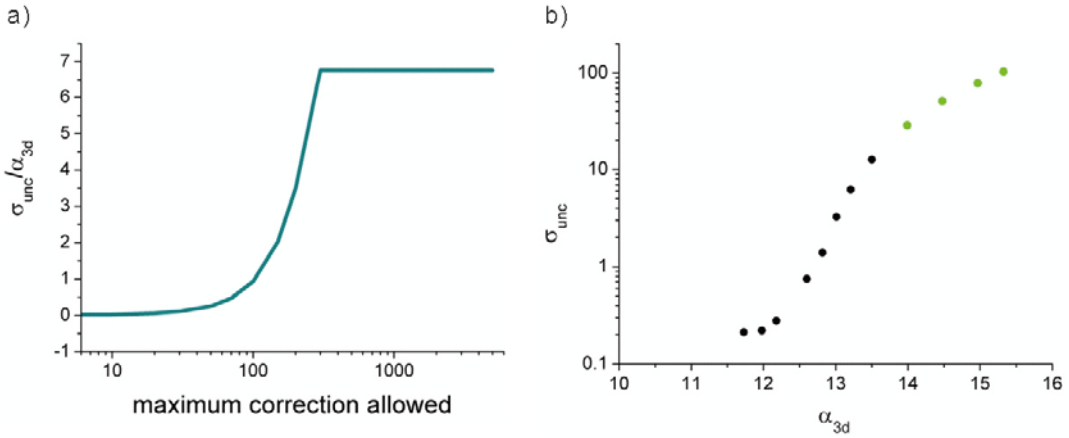


Figure 4-4. *Orientation-bias correction threshold values and consequences on density and standard deviation, illustration of KLX09C_FSM_N: a)* σ_{unc}/α_{3d} *= f (threshold value), the maximum value; and b) evolution of* σ_{unc} *versus* α_{3d}*..*

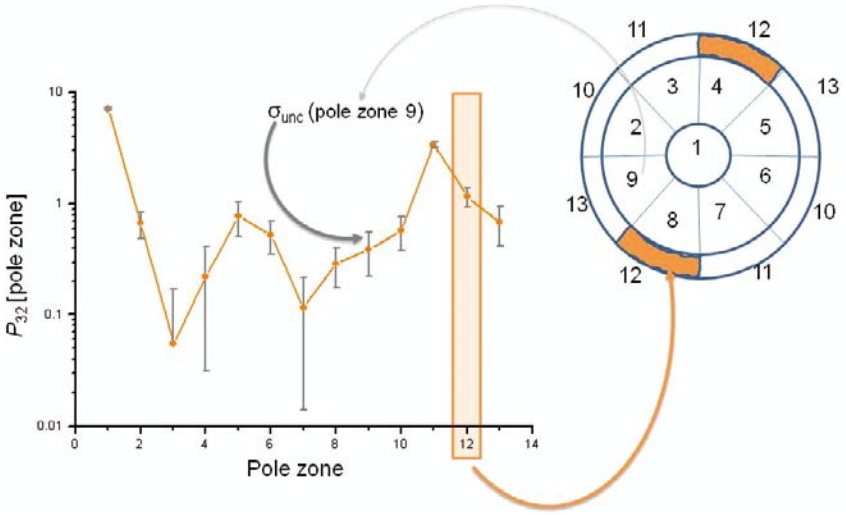


Figure 4-5. *Parent density distribution [$\alpha_{3d}(\theta,\varphi)$; $\sigma_{unc}(\theta,\varphi)$] representation expressed according to P_{32}.*

Figure 4-6, we provide the equivalent density distributions for several sections of borehole KFM01C. The subdivision is performed according to FD and DZ divisions (Figure 4-6*a*), leading to four potential distributions representing *KFM01C-FFM02a*, *-FFM02b*, *-FFM02 affected by DZ* and *-FFM01 affected by DZ* (borehole datasets subdivisions and DZ exclusion are explained in Section 3). As expected, the grid pole zones with the highest densities (pole zones 11 and 1 for FFM02b) reflect the highest concentrations observed in a contoured stereonet representation (Figure 4-6*b*, pink curve and contoured stereonet in Figure 4-6*c*).

We emphasize that the x-y diagram representation not only allows a relative intensity comparison between various pole zones, but also an absolute comparison in density. This enables direct density/orientation comparison between datasets.

Finally, we show how the total number of fractures in a dataset is handled. Even if no fracture pole falls within a grid pole zone, the contribution of some poles in the surrounding domains can be different from zero, because of measurement precision associated with each datum. This is the case in pole zones 2 and 3 for KFM01C-FFM02a. However, due to the uncertainty/precision value linked to each datum, the final contribution into these pole zones is not null, though it is quite low and associated with large uncertainty bars. Conversely, dataset *FFM01 affected by DZ* apparently displays a lower ratio between uncertainty and density.

The computation of couples $\{\alpha_{3d}(\theta,\varphi)\ ;\ \sigma_{unc}(\theta,\varphi)\}$ over each dataset of the Forsmark and Laxemar site is performed next, and the results are analyzed in Section 4.6.

4.5 Variability characterization

4.5.1 Definition

In this section, **variability** is introduced into the modeling process.

If uncertainty aspects are discarded in this part, parent density can still be computed for any dataset. However, this density is associated with a specific location and a finite-size sample. In order to reflect the underlying DFN stochastic model, this measure must be accompanied with a variability assessment. This provides a framework for quantifying the variability scaling through the characterization of the gap between local density variations and a single mean value associated with the sample size (Figure 4-7). Even a Poisson process results in a natural variability (see below). Two realizations of a single Poisson process display some variations due to the stochastic process. The resulting variability must be quantified and considered when comparing the two realizations; otherwise, one compares particular realizations rather than underlying stochastic processes (and models).

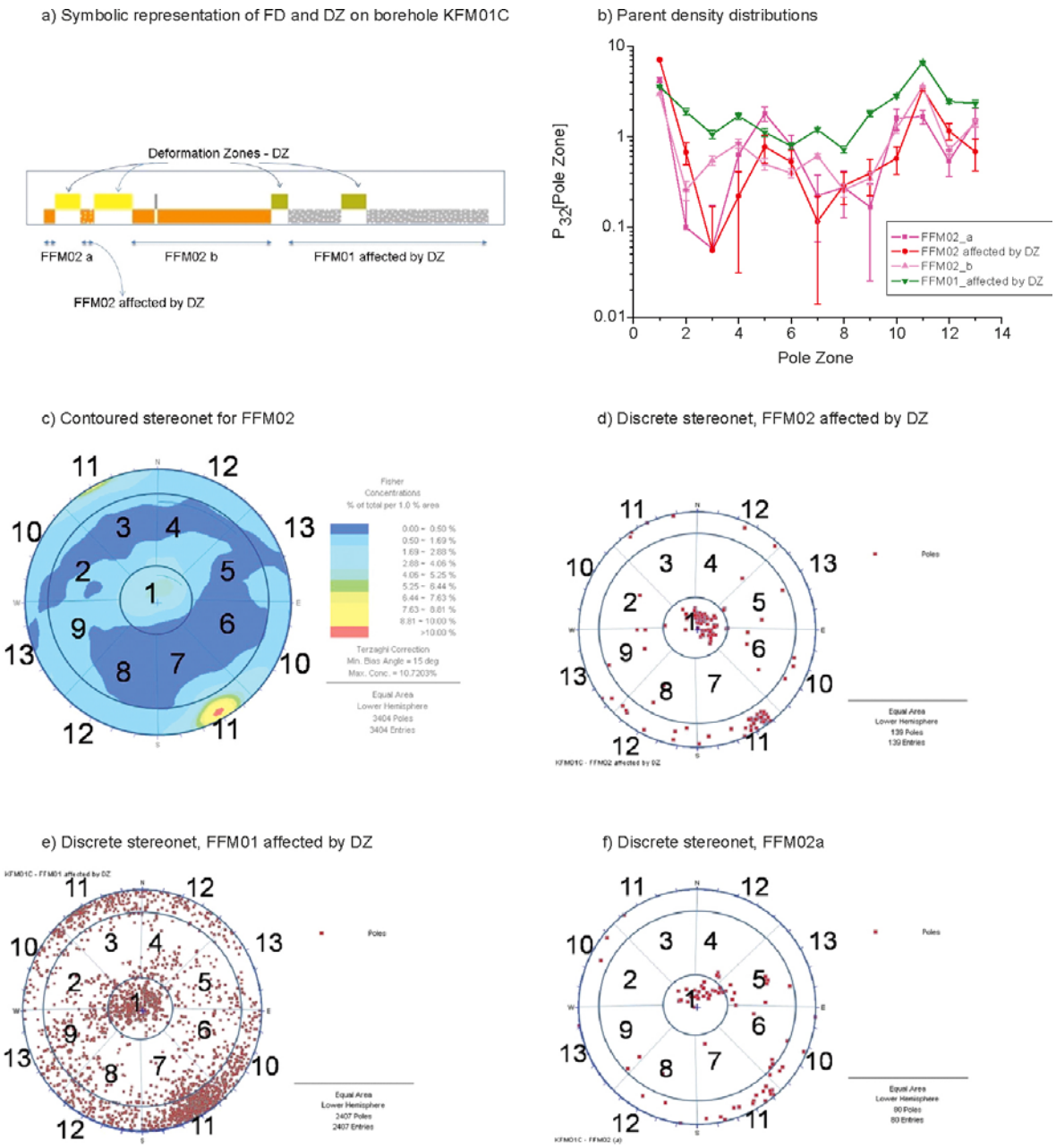


Figure 4-6. *Borehole KFM01C: a) symbolic representation, b) density distributions for the various subsets identified in a), c) contoured stereonet for all subsets of FFM02 and d) to f), discrete stereonets for respectively, domain "FFM02 affected by DZ", "FFM01 affected by DZ" and FFM02a.*

The DFN model parameters thus determined can be used to extrapolate/interpolate values between the available finite-size samples. Therefore, a variability analysis is necessary to determine the range of validity of a density assessment or to link finite-size sample measurements to models.

In practice, it is not possible to measure directly the variability at the sample size: indeed, the whole sample is equivalent to a unique realization, preventing any assessment of variability. Therefore, the model **variability**, noted $\boldsymbol{\sigma_{var}(L)}$, arises from a scale analysis (Figure 4-7).

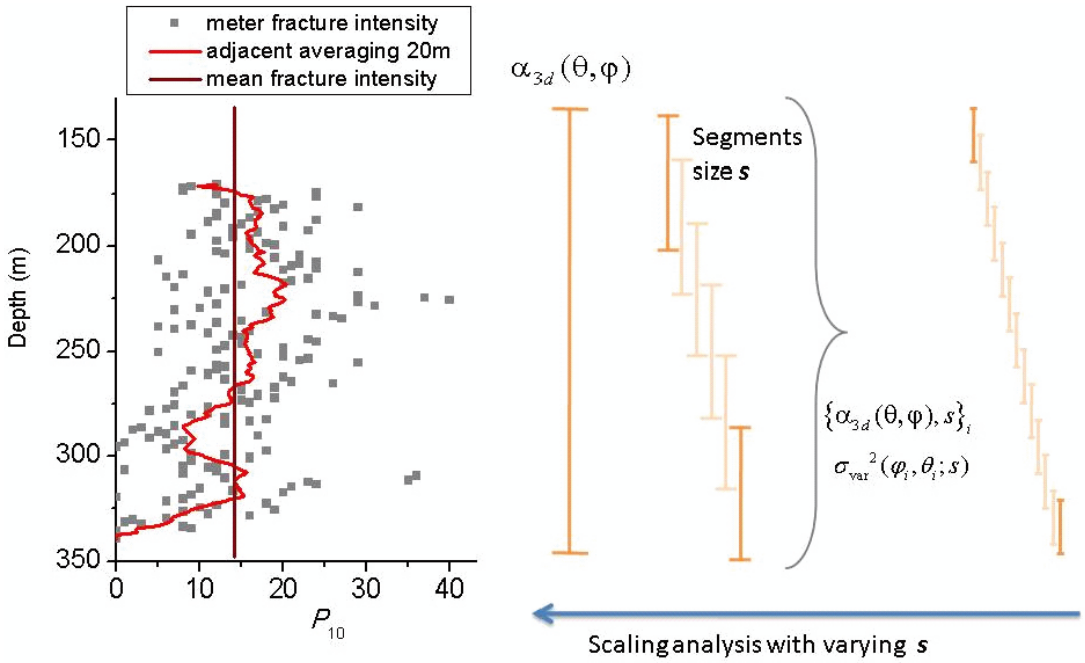


Figure 4-7. *Principle of the variability scaling analysis (see text for details).*

In order to quantify the DFN natural variability, we consider the evolution, with respect to the scale of observation (s), of the fracture-density-distribution standard deviation, $\sigma_{var}(\theta, \varphi;s)$:

$$\sigma_{\text{var}}(\varphi,\theta;s)=\sqrt{\overline{\left(\alpha_{3d}(\varphi,\theta;s)\right)^2}-\overline{\left(\alpha_{3d}(\varphi,\theta;s)\right)}^2} \qquad 4\text{-}14$$

where

$$\overline{\alpha_{3d}(\varphi,\theta;s)}=\frac{1}{N_w}\sum_{m=1}^{N_w}\left[\alpha_{3d,m}(\varphi,\theta;s)\right]$$

and

$$\overline{\left(\alpha_{\alpha_{3d}}(\varphi,\theta;s)\right)^2}=\frac{1}{N_w}\sum_{m=1}^{N_w}\left(\alpha_{3d,m}(\varphi,\theta;s)\right)^2$$

$\sigma_{var}(\theta, \varphi;s)$ is the standard deviation associated with the fracture density distribution, $\alpha_{3d}(\theta, \varphi;s)$, calculated over sub-samples (or windows) of size s defined within a domain of size L. N_w is the number of windows of size s defined over the whole domain.

We illustrate the variability analysis through the simultaneous comparison of a "real" dataset ("KFM01C-FFM01 affected by DZ") and its "Poisson equivalent", which can be seen as a modified dataset in which all the fracture intercept positions are redistributed over the sample (borehole section), according to a Poisson process. By doing so, any specific spatial correlation related to the fracturing organization existing in the original dataset is removed. Both datasets are plotted in Figure 4-8*b*, where the fracture frequency by meter, P_{10}, is plotted as a function of relative depth. Qualitatively, the Poisson equivalent appears more regular than the original dataset. This discrepancy is reflected in the behavior of $\boldsymbol{\sigma_{var}(s)}$, provided in Figure 4-8*a*.

$\boldsymbol{\sigma_{var}(s)}$ decreases with increasing scale: the larger the window of observation, the smaller the resulting model uncertainty due to variability. The trend follows a power-law model whose behavior is masked by finite size effects. The Poisson case displays the power-law trend, with an exponent equal to –0.5. The exponent is smaller for the real data case, leading to the largest values of $\boldsymbol{\sigma_{var}(s)}$ for any scale of observation.

Because the standard deviation associated with α_{3d} cannot be measured directly, $\boldsymbol{\sigma_{var}(L)}$ is extrapolated to the dataset size (size L of a borehole section), from the scaling evolution of $\boldsymbol{\sigma_{var}(s)}$. In the case presented in Figure 4-8, we observe , at the sample scale, a standard deviation of the real dataset more than 4 times larger than what is expected from the Poisson dataset. This underlines the difficulty in assessing the parent DFN density when spatial correlations are large.

Finally, thus assessed, $\sigma_{var}(L)$ is the standard deviation associated with density α_{3d}, relevant at the sample size. In other words, it represents the uncertainty on the computed value of α_{3d} due to variability. The deduced parent distribution model is $\alpha_{3d}(\theta, \varphi) \pm \sigma_{var}(\theta, \varphi)$.

In summary, the variability analysis thus entailed:

a) is a way to quantify spatial correlation of the fracture system (quantifying its departure from Poisson processes is a way to define the DFN degree of spatial correlation),

b) illustrates the role of sampling conditions (sample statistical robustness, sample size) on the precision of the resulting parameter estimation ($\boldsymbol{\sigma_{var}}$ decreases while the observation scale increases); and,

c) provides quantitative assessment of the resulting model uncertainty: $\boldsymbol{\alpha_{3d}(\theta, \varphi) \pm \sigma_{var}(\theta, \varphi)}$.

4.6 Uncertainty and variability comparison

Both the uncertainty ($\boldsymbol{\sigma_{unc}(\theta,\varphi)}$) and the variability ($\boldsymbol{\sigma_{var}(\theta, \varphi)}$) concepts introduced above result in the definition of a range of confidence around the parent density distribution, or stochastic standard deviation: $\boldsymbol{\alpha_{3d}(\theta,\varphi) \pm \max(\sigma_{unc}(\theta,\varphi), \sigma_{var}(\theta, \varphi))}$. This **range of confidence** also is called a **model uncertainty**. In the following we compare the density, uncertainty and variability, in order to assess which of them supersede the others.

Investigations are conducted globally: $\boldsymbol{\alpha_{3d}(\theta, \varphi)}$, $\boldsymbol{\sigma_{unc}(\theta,\varphi)}$ and $\boldsymbol{\sigma_{var}(\theta, \varphi)}$ are computed for all the available datasets at Forsmark and Laxemar (see Section 3 for the datasets definition). All computations and result presentations are adapted to provide $\boldsymbol{P_{32}(\theta, \varphi)}$, instead of $\boldsymbol{\alpha_{3d}(\theta, \varphi)}$ (see Equ. 4-5).

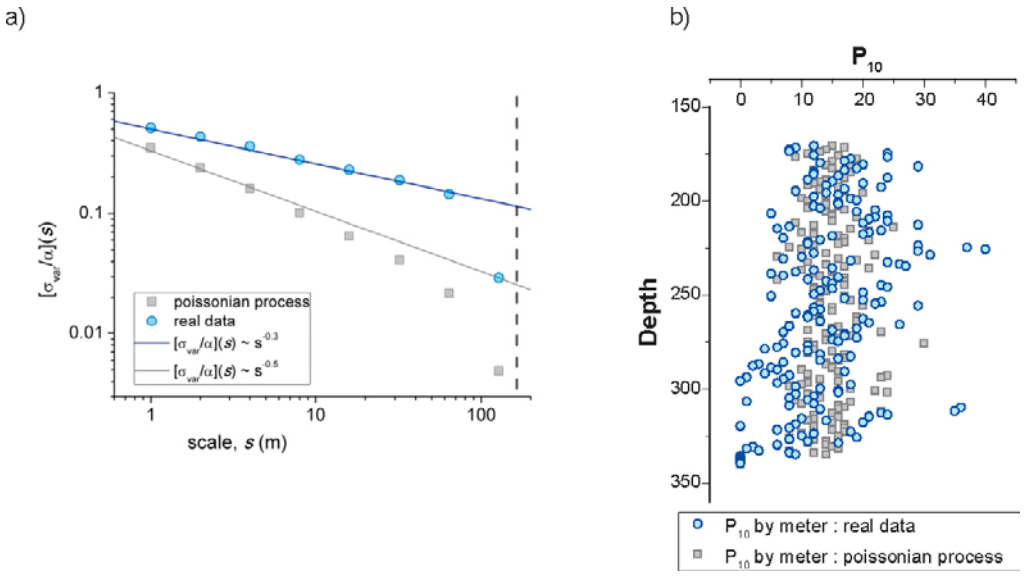


Figure 4-8. *Variability analysis illustrated for dataset KFM01C-FFM01 affected by DZ: a) Evolution of $\sigma_{var}(s)/\alpha_{3d}$, comparison between a synthetic equivalent Poisson dataset (grey symbols and grey line) and real data (blue symbols and blue line). The sample/section size is denoted as a dashed line ($\boldsymbol{L}$ = 164). **b)** Apparent fracture intensity by meter, $\boldsymbol{P_{10}}$, for the Poisson dataset (grey symbols) and the real dataset (blue symbols).*

Results (e.g. in Figure 4-9) are presented as clusters of points in which each point corresponds to a computed couple of $\{\boldsymbol{P_{32}(\theta,\varphi)};\boldsymbol{\sigma_{unc}(\theta,\varphi)}\}$ or $\{P_{32}(\theta,\varphi);\sigma_{var}(\theta,\varphi)\}$. **For each dataset,** several levels of refinement (regarding the angular or pole zones definition, see Figure 4-1) are investigated. Thus, the resulting clusters of points reflect either one angular division or the other. Note that density from both grids are related through:

$$\alpha_{3d} = \sum_{\text{zone }(\theta,\varphi)=1}^{13} \left[\alpha_{3d}(\theta,\varphi)\right]$$

On average, densities from the *13-pole-zones grid* are 13 times smaller than global densities (on the *1-pole-zone grid*).

At the Forsmark site, shown in Figure 4-9*a* and *c*, we observe the following:

- The computed $\boldsymbol{\sigma_{unc}(\theta,\varphi)/P_{32}(\theta,\varphi)}$ values (*empty brown squares*) spread over a global trend defined by:

$$\left.\frac{\sigma_{unc}}{P_{32}}\right|_{(\theta,\varphi)} \propto P_{32}^{-0.5}$$

- The trend thus defined (*brown straight line*) from the *13-pole-zones grid* (i.e. for very small density values), is consistent with values of $\boldsymbol{\sigma_{unc}/P_{32}}$ computed for larger densities (*brown circles filled in blue*).

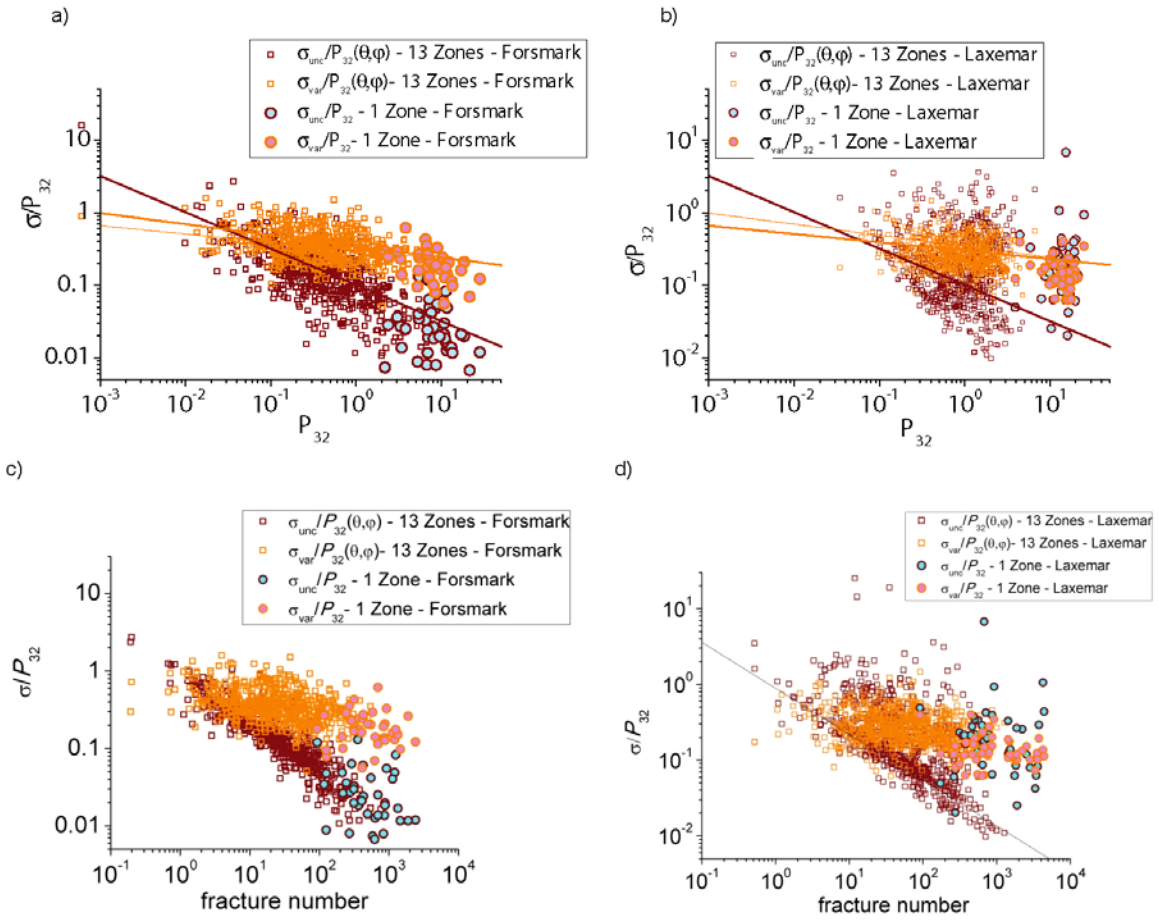


Figure 4-9. *Variations of $[\sigma_{unc}/P_{32}]$ and $[\sigma_{var}/P_{32}]$ at Forsmark (a and c) and Laxemar (b and d) for densities computed from the 13-pole-zones grid (θ,φ) and for all orientations together. The variation is presented as a function of density P_{32} in* ***a)*** *and* ***b)****, and as function of the fracture number in the dataset in* ***c)*** *and* ***d)****.*

- In a similar manner, the ratio $\boldsymbol{\sigma_{var}(\theta,\varphi)/P_{32}(\theta,\varphi)}$ (*empty orange squares*) decreases with increasing $\boldsymbol{P_{32}(\theta,\varphi)}$, with a global trend defined by:

$$\left.\frac{\sigma_{\text{var}}}{P_{32}}\right|_{(\theta,\varphi)} \propto P_{32}^{-0.15}$$

- The trend thus defined is consistent for values of $\boldsymbol{\sigma_{var}/P_{32}}$ computed for larger densities (*pink-filled orange circles*).
- The two trends intersect for a density P_{32} close to 0.03. Thus, for datasets related to very small densities, the uncertainty due to data precision is predominant, whereas variability dominates for larger densities.
- In any case, most computed values remain below a threshold, such that.

$$\frac{\sigma_{unc}}{P_{32}} < 1 \text{ and } \frac{\sigma_{\text{var}}}{P_{32}} < 1$$

This means that the range of confidence established for the density distribution values is smaller than the density value itself. In other words, density values are not completely swamped into the overall uncertainty and variability, but, rather, have significance.

- Finally, we try to determine what can control the observed uncertainty and variability. The number of fractures in a dataset is found to be a critical element for the uncertainty aspect (Figure 4-9*c*): indeed, the cluster of points $\boldsymbol{\sigma_{unc}(\theta,\varphi)/P_{32}(\theta,\varphi)}$ is tightened around the power-law trend.

Finally, we check that the obtained results have a general significance and need not to be refined according to site specifics; thus, the Laxemar database is analyzed according to the same procedure, and results are displayed in Figure 4-9*b* and *d*.

Densities computed from the borehole data at Laxemar are more compact than those at Forsmark: Laxemar densities have few values of P_{32} in the range $[10^{-2}; 10^{-1}]$, and its global densities are mostly larger than 10 (whereas they are below 10 at Forsmark).

We observe the following.

- The variability scaling behavior is very similar at both sites.
- Uncertainty values at Laxemar display a more complex behavior: the cluster of points can be divided into two parts: part of the data follow a trend equivalent to that observed at Forsmark (with a decrease in $(\sigma_{unc}/P_{32}) \sim (P_{32})^{-0.5}$), and the rest falls in the same range as the variability data. Moreover, the values computed from the whole dataset (*1-pole-zone grid*) reflect this second behavior more frequently. At this stage, we do not have a complete explanation for this double behavior; further investigation is needed to solve this issue.
- In any case, most computed values are below a threshold such that.

$$\frac{\sigma_{unc}}{P_{32}} < 1 \text{ and } \frac{\sigma_{\text{var}}}{P_{32}} < 1$$

4.7 Conclusions on the variability and uncertainty

The uncertainty and the variability concepts are quantitatively introduced into the modeling process. This allows determination of a range of confidence around the parent density distribution and, consequently, assessment of the aspects that are critical for the modeling process.

The process is applied to Forsmark and Laxemar sites. Analyses are performed globally and are based on all the relevant available data resource (51 datasets defined at Laxemar and 60 at Forsmark, according to borehole and FD references).

We conclude that:

- Variability $\sigma_{var}(\theta,\varphi)$ is predominant over uncertainty $\sigma_{unc}(\theta,\varphi)$.
- Uncertainties $\sigma_{unc}(\theta,\varphi)$ at Forsmark correlate strongly with the number of elements in the datasets. The larger the number of elements, the smaller the remaining uncertainties. However, some values at Laxemar do not fit well with this behavior. In the framework of the present study, we did not have the time to unravel this complex behavior.
- Overall uncertainty and variability do not mask information completely. Therefore, it is possible to determine significant values for the DFN model parameters.

Conditions, then, are fulfilled such that a systematic comparison between parent density distributions at the site scale can be conducted.

5 Statistical fracture domain analysis

5.1 Introduction

This section presents the last steps of the process illustrated in Figure 1-1. An automated sorting process is developed in order to get the capacity to quantitatively assert whether two local datasets (or two sets of DFN parameters) are equal or different. More generally, the sorting process treats simultaneously any number of initial objects (which are here the parent distributions – $\alpha_{3d}(\theta,\varphi) \pm \sigma_{3d}(\theta,\varphi)$ – built in previous sections). Beyond the methodological development, the objective is therefore to sketch out the DFN site model.

Let recall that the elementary bricks (datasets) of the process arise from the local divisions of boreholes according to Fracture Domains (section 3). However, the link between FDs from one location to the other is ignored and sorting is based exclusively on the datasets statistical parameters. The local datasets (e.g. set KLX09-FSM_N) are the natural candidates for initiating the site DFN consistency analysis. Indeed this initial division partly arises from the available geological knowledge of the site and other qualitative information and from apparent fracture statistics. The sorting process is next fully quantitative. Systematic computation of parent density distributions, $\alpha_{3d}(\theta,\varphi) \pm \sigma_{3d}(\theta,\varphi)$, associated to each dataset, is a prerequisite.

Because the sorting process is based on parent distributions of densities, it is adapted to include any initial datasets, whatever their orientations (for the boreholes) or spatial location throughout the site. The outcome of the sorting process consists of grouping the datasets into several statistical classes whose equivalent density distributions are also computed. Within one class, all elements are considered to belong to the same parent domain, with fracturing properties satisfactorily described by the mean properties assigned to the class itself. The final result stems from a combination of the two extreme and opposite trends, as described below:

- The number of classes is equal to the initial number of units introduced in the classification process – in other words, none of the defined units has fracturing properties that can be considered as equivalent to another; in such a case, variability throughout the site is extremely high and cannot be encompassed locally by the DFN model used.
- All candidate units are gathered into one single class – in other words, the fracturing properties are rather homogeneous, and a single DFN model set of parameters should be enough to describe fracturing statistically.

We underline also that, within the sorting process, the quantitative comparison between fracturing properties can be undertaken through various levels of refinement. The method may reflects simply the sensitivity to global densities or more precisely the sensitivity to orientation distribution shape. Besides, the relevancy of initial units defined at the borehole local scale is critical (issue 7 in Introduction).

The sorting method developed here currently is under progress, and will be submitted for scientific publication as "An automatic comparison method of probability density functions with uncertainties**"** /Le Goc et al. in prep/**.** Its principles and mathematical framework are provided in Sections 5.2.1 and 5.2.2.

5.2 Method

5.2.1 Bibliographic review

Observation of natural processes can be described probabilistically using probability density functions (PDF). A PDF represents the density distribution of a physical quantity over a sample space. This physical quantity can be observed for different objects. The comparison of different objects based on their PDF is not a trivial task. Thus, an objective method is required in order to systematically find similarities and differences between objects. Such methods are used in a wide range of scientific fields, such as biology and biochemistry /Cheng et al. 2006, Datta and Datta 2006/, mathematics /Garai and Chaudhuri 2004/, communication /Kim and Yamashita 2007/, climate research /Vrac et al. 2007/, soil analysis /Sarbu et al. 2007/ and rock mechanics /Hammah and Curran 1998, Hofrichter

and Winkler 2006/. The method proposed in /Le Goc et al. in prep/, is relevant for all these fields, but is applied here only to fracture properties description in natural media. The mathematical framework is recalled below.

5.2.2 Mathematical Framework

This is a classification method: it generates classes with specific characteristics in partitioning a space into different subspaces. It is inspired by the agglomerative hierarchical clustering algorithm /Chelcea et al. 2006/, and generates disjointed classes to emphasize object groups with same properties.

Let E be an object set containing n objects (our objects are the local datasets mentioned above). We consider collections of non-void subsets of E called classes. (For further details, see /Gordon 1999/.) If C is a class collection and X is an element of C, C is a valid division of E if C contains all the subset of E and for every X and Y of C, $X \cap Y = \varnothing$. The principle of the agglomerative hierarchical clustering is, from a class collection containing all singletons, to aggregate iteratively two classes until the largest degree of heterogeneity of element X of C reaches a certain value. Moreover, two classes, X and Y, are aggregated only if they are the nearest in the set. When two classes are aggregated, the distance between them, written as $d(X,Y)$, represents the degree of heterogeneity of $Z = X \cup Y$.

The algorithm contains the steps shown in Table 5-1.

At the end, we obtain a collection of classes, C, where all classes can be considered as different (with respect to a treshold d_{max}) from the others, and where all elements of a class can be considered as equivalent (with the same characteristic).

5.2.3 Application to fracture fields

5.2.3.1 Basic definitions

In the sorting algorithm framework, a **distribution** refers simply to a parent density distribution:

$$\{\ \alpha_{3d}(\theta,\varphi) \pm \mathrm{Max}\ (\sigma_{var}(\theta,\varphi)\ ,\ \sigma_{unc}(\theta,\varphi))\ \}_{(\theta,\varphi)\in\ldots}$$

defined in Equ. 4-3, or, in an equivalent manner, in Equ. 4-5 for $P_{32}(\theta,\varphi)$, relative to a dataset or a sample. One **element** of one distribution is noted $(\alpha_{3d};\sigma)_j^i$, where i refers to the distribution and j to the range (θ,φ). The latter notation refers to the grid pole zones defined in Section 4.2.2 (Figure 4-1).

Table 5-1: Sorting process algorithm.

(0) Set $i = 0$ and define M_0 containing all the singletons of E.

(1) Set $i = i+1$ and merge the two classes X_i and Y_i such as $\forall \{X,Y\} \subseteq M_{i-1}, d(X_i,Y_i) \leq d(X,Y).$

(2) Set $M_i = M_{i-1} \cup \{X_i \cup Y_i\} - \{X_i\} - \{Y_i\}$

(3) Optimization of classes:
$\forall \{X,Y\} \subseteq M_i, \forall x \in X, \bar{x} = \mathrm{mean}(X), \bar{y} = \mathrm{mean}(Y+x), \text{if } \max(d(x,\bar{x})) > \max(d(x,\bar{y}))$
set $X = X \setminus x$ and $X = X \setminus x$.

(4) Repeat (3) while it is effective.

(5) Repeat (1) to (4) until $\forall \{X_i,Y_i\} \subseteq M_i, \max(d(X_i,Y_i)) >= d_{max}$, with d_{max} a fixed criterion.

A **class** arises from the sorting process (Table 5-1). It is defined as a set of compatible distributions, knowing that compatibility is defined with respect to a threshold (d_{max}). It is denoted as $\{\alpha_{3d}^{i},\sigma^{i};i=1\cdots N_{cs}\}$, where N_{cs} is the number of distributions in the class *CS*. A mean density and corresponding standard deviation (Equ. 5-1 and Equ. 5-2) are associated with a class. These are computed from the distributions included in a class in the following manner. With N_{CS} the number of distributions in a class (referred as Class *CS*) and *j* is the *j-th* point of the distribution, the mean density distribution (α_{CS}) is built by averaging the density values inversely weighted by their standard deviations:

$$\left(\alpha_{CS}\right)_j = \frac{\sum_{i=1}^{N_{CS}} \frac{\left(\alpha_{3d}\right)_j^i}{\sigma_j^i}}{\sum_{i=1}^{N_{CS}} \frac{1}{\sigma_j^i}} \qquad 5\text{-}1$$

In addition, the standard deviation around the mean value ($\boldsymbol{\sigma}_{CS}$) is defined by:

$$\left(\frac{1}{\sigma_{cs}}\right)_j = \frac{1}{N_{CS}} \sum_{i=1}^{N_{CS}} \frac{1}{\sigma_j^i} \qquad 5\text{-}2$$

The above definitions guarantee the conservation of the final properties of a class that would encompass all the distributions: from any stage of the sorting process, the mean distribution resulting from the merging of all distributions ends up to a unique global distribution, $(\alpha_{mean})_{TOT}$ in:

$$\left(\alpha_{mean}\right)_{TOT,j} = \frac{\sum_{i=1}^{N_{CS}} \frac{\left(\alpha_{3d}\right)_j^i}{\sigma_j^i} + \sum_{i=N_{CS}+1}^{N} \frac{\left(\alpha_{3d}\right)_j^i}{\sigma_j^i}}{\sum_{i=1}^{N_{CS}} \frac{1}{\sigma_j^i} + \sum_{i=N_{CS}+1}^{N} \frac{1}{\sigma_j^i}} = \frac{N_{CS} \cdot \alpha_{3d,C} + \sum_{i=N_{CS}+1}^{N} \frac{\left(\alpha_{3d}\right)_j^i}{\sigma_j^i}}{N_{CS} \cdot \frac{1}{\sigma_c} + \sum_{i=N_{CS}+1}^{N} \frac{1}{\sigma_j^i}} \qquad 5\text{-}3$$

In the above equation, each distribution contributes either directly, as long as it is not included in a class, or through the mean class parameters when it is included into one (say, Class *C*S in Equ. 5-3). Accordingly, the sorting process defines classes whose equivalent parameters (mean and standard deviation) have the following properties:

- Within a class, the weight attributed to each distribution, in order to define the mean density, is inversely proportional to the standard deviation. In other words, distributions with relatively small standard deviations are more significant; conversely, the contribution of a distribution with very large standard deviations is insignificant in defining the mean trend of a class.
- The standard deviation associated with a class arises from the harmonic mean of the standard deviations of the density distribution. Accordingly, it tends to reflect the smallest standard deviations.

As a consequence, the more distributions are included in a class, the more precise is the resulting definitions of the class parameters.

5.2.3.2 Definition of distance

The **distance** between two distributions fundamentally is defined to reflect their probability of belonging to the same parent distribution. It requires that both mean and standard deviations of a distribution be defined. It is quantified from the difference between density values, normalized by their related standard deviations. First, for two distributions *m* and *n*, the distance is defined by:

$$d\left(m,n\right) = \frac{1}{N_Z} \sum_{j=1}^{N_Z} \frac{\left|\left(\alpha_{3d}\right)_j^m - \left(\alpha_{3d}\right)_j^n\right|}{\sigma_j^m + \sigma_j^n} \qquad 5\text{-}4$$

where N_Z refers to the grid pole division defined in Section 4.2.2 (Figure 4-1). $N_Z = 1$ for the *1-pole-zone grid*, $N_Z = 9$ for the *dip grid*, and $N_Z = 13$ for the remaining one.

Thus defined, distance can be computed between two basic distributions, or between a distribution and the equivalent distribution associated to a class. It also is used to define the intrinsic distance, or dispersion, associated to a class (Equ. 5-5).

The distance between distributions has the following meanings:

- In a simple case ($N_Z = 1$), the distance between distributions larger than 1 reflects a lack of overlapping between the standard deviations, as happens for distributions *A* and *C* in the example shown in Figure 5-2*b*. Thus, the difference between two density values is larger than the sum of their related standard deviations.
- In contrast, if the distance between distributions is smaller than 1, there is overlapping between the two distributions.
- When $N_Z > 1$, the distance is taken simply as the average from each point of the density distributions; accordingly, this reflects a trend and does not prevent from any local discrepancy. Locally, a discrepancy between two points can be compensated by another match between other points of the distributions. This is related to the difficulty of defining the gap between two distributions, or two set of discrete points. We could have chosen a more restrictive criterion by applying it point-by-point rather than on average.
- An increase in standard deviation leads to a relative decrease of distances, thus bringing distributions closer.

Finally, an internal distance, $\boldsymbol{d_{\text{class}}}$, is associated to each **class.** It reflects the class internal cohesion, and it is defined as the maximum distance between the class mean distribution (α_{CS}) and its constitutive elements (distributions $n \in CS$) . It is noted d_{CS} in the following:

$$d_{CS} \equiv d(\text{class } CS) = \max\left\{d\left(\alpha_{CS}, \alpha_n\right); n \in CS\right\} \qquad 5\text{-}5$$

In the sorting process, we consider the maximum class distance among the classes after each step (Equ. 5-6) and the class aggregation is pursued until it reaches a threshold.

$$\max\left[d_{CS}\right]_C = \max\left(d_{CS}; \{CS \in C\}\right) \qquad 5\text{-}6$$

5.2.3.3 Sorting procedure and critical threshold

During the sorting process (Section 5.2.2), the class building procedure is iterated as long as the most-dispersed class displays an internal distance, d_{CS}, smaller than a threshold value d_{max}. Distributions are aggregated as long as the criterion is fulfilled.

The critical threshold reflects the distance below (or above) which two distributions (or two classes) are set similar (or not):

- If the distance between two classes, CS_1 and CS_2, d (CS_1, CS_2), is very small (if the gap between values is much smaller than the associated uncertainties), then both classes are statistically undistinguishable (at the process initial step, classes are made of single distributions).
- Conversely, if the distance is very large, both classes are statistically distinct.

In practice, having a *distance* (Equ. 5-4) smaller or equal to 1 reflects that standard deviations do overlap. Thus, fixing a threshold below 1 is more restrictive, and, in contrast, accepting a larger *distance* tends to aggregate distributions according to a relative closeness (within the total set of data), whereas they are not really similar strictly speaking. In other words, accepting a distance threshold d_{max} larger than 1 results in artificially increasing the standard deviations of the distribution.

5.2.3.4 Coherence index[2]

Ideally, the resulting classification contains classes that do not overlap. In reality, however, some distributions might be associated with two or more classes. An efficient validity indicator of the classification is required to compute, for each distribution, its probability of belonging to any class. In other words, distributions and sorting process are initially used to define classes with statistical properties expected to represent general trends over the site. Next, the probability of each distribution to belong to each of the defined classes is considered. Accordingly, the coherence index is based on the probability, for a distribution, of belonging to a class. It is expressed as follows:

$$p\left(x \in X \; Y\right) = \frac{d^*\left(x, \overline{y}\;\right)^{-1}}{\sum_{i}^{N_{\text{class}}} d^*\left(x, \overline{y}_i\right)^{-1}} \qquad \text{5-7}$$

$$\text{with} \qquad d^*\left(x, \overline{y}_i\right)^{-1} = \begin{cases} d\left(x, \overline{y}_i\right)^{-1} & \text{if } d\left(x, \overline{y}_i\right) < d_{\max} \\ 0 & \text{else} \end{cases}$$

$d^*(x,\overline{y}_i)$ is the distance between a distribution x and a Class Y_i denoted by its mean distribution $\overline{y}_i$. C refers to the whole set of classes (collection). These probabilities are meant to evaluate the classification.

We finally compute a coherence index IC for each class $X \epsilon\ C$ with the formula:

$$IC(X \text{ on } X) = IC(X) = \frac{\sum_{i}^{N_X} P\left(x \in X, X\right)}{N_X} \qquad \text{5-8}$$

$$IC(X \text{ on } Y) = \frac{\sum_{i}^{N_X} P\left(x \in X, Y\right)}{N_X}$$

The first index reflects how the distributions from one class do overlap into others, independent of the number of distributions in the class. In complement, the following IC_w index can be computed:

$$IC_w(X) = \sum_{i}^{N_X} p\left(x \in X, X\right) \qquad \text{5-9}$$

$$IC_w(X \text{ on } Y) = \sum_{i}^{N_X} p\left(x \in X, Y\right)$$

This second parameter provides a complementary criterion to characterize overlapping, in which the initial weight (or number of distributions) associated to each class is taken into account.

Once a classification is performed, the class robustness is assessed further through computation of these indexes. The real robustness of a class arises from the comparison between IC from its own class and from its "projection" into other classes.

In other words, the coherence index is defined from the above-defined probabilities of a set of distributions grouped in a class with regard to any other class. Therefore, the coherence index represents both the internal relevancy of one class as well as the overlapping of one class into another.

[2] from /Le Goc et al. in prep/.

5.3 Practical use of the method

5.3.1 Input data: dataset density distributions

In the sorting and class-identification processes, the Max **($\sigma_{var}(\theta,\varphi)$, $\sigma_{unc}(\theta,\varphi)$)** is kept to represent the DFN-model density standard deviation (or error bar). As explained in Section 4, the variability parameter usually dominates the uncertainty parameter (Figure 5-1).

5.3.2 Sorting process illustration

The sorting method aims at defining statistically relevant classes from an initial set of density distributions. It aims to distinguish, up to a certain threshold, the closeness between distributions. Thus, the final sorting results from the combination of mean densities, associated standard deviations and the threshold allowed for the distances.

Consider a simple case with three basic density distributions, *A*, *B* and *C*. Each is defined by a unique value (and standard deviation), as it is for the *1-pole-zone grid* type. Both examples, illustrated in Figure 5-2*a* and *b*, differ only through the associated standard deviations: $\sigma = 2$ for Case *a* and 1 in Case *b*. The sorting process is set with $d_{max} = 1$. In the first case ($\sigma_A = \sigma = \sigma_C = 2$), the sorting process creates one single class with the three distributions; in the second case (Figure 5-2*b*), the process ends up with one class encompassing *B* and *C* (*A* is isolated).

More precisely, one-to-one distances between the three distributions initially are computed (Equ. 5-4). In the example, they range as:

$$d(B,C) < d(A,C) < 1 < d(A,B)$$

In the first case ($\sigma = 2$, Figure 5-2*a*), the two closest distributions, *B* and *C*, are grouped initially into one class. Next, because the distance between the class thus formed and *A* is still below the threshold, *A*, *B* and *C* finally fall into one single class.

In the second case, the process ends sooner: the distance between *A* and the class that includes *B* and *C* is larger than the threshold allowed. Under these conditions, *B* and *C* belong to a single class, and A is isolated. In addition, the three distributions would fall into a single class if the threshold d_{max} were raised sufficiently. In general, increasing the threshold can be viewed as a systematic increase of the standard deviations associated with all density distribution values.

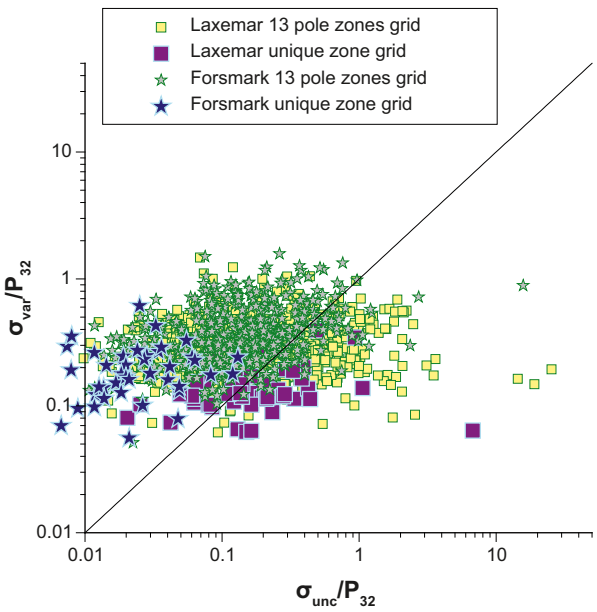


Figure 5-1. *$\sigma_{var}(\theta,\varphi)/P_{32}$ versus $\sigma_{unc}(\theta,\varphi)/P_{32}$ for all couples computed at Laxemar and Forsmark. Max ($\sigma_{var}(\theta,\varphi)$, $\sigma_{unc}(\theta,\varphi)$) is kept in the sorting and class-identification processes.*

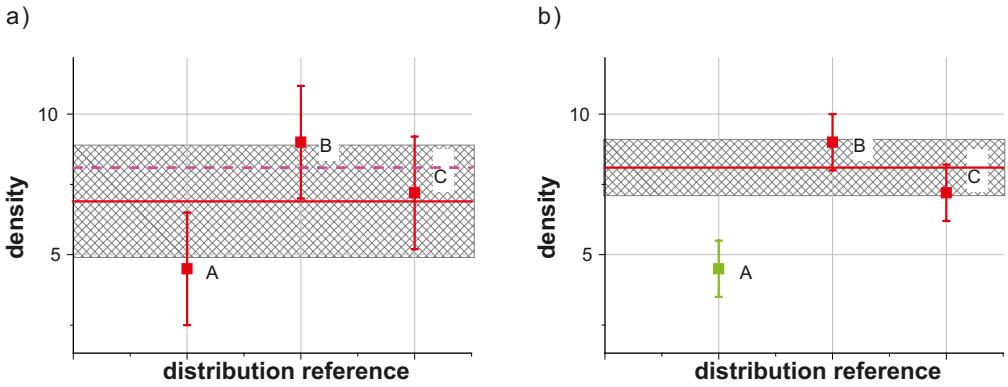


Figure 5-2. *Distance calculation and distribution merging into classes. Illustration for three initial distributions **A**, **B** and **C**. The final class density is shown by the red line, and the associated standard deviation is shown by the sparse gridded box: a) $\sigma_A = \sigma_B = \sigma_C = 2$, all distributions are merged into one class (the first step of sorting-process groups **B** and **C**, with a mean symbolized by the magenta dashed line); b) $\sigma_A = \sigma_B = \sigma_C = 1$ (only distributions **B** and **C** are merged into one class, and distribution A is isolated).*

The sorting process and threshold mechanism are illustrated here with a very simple case. In practice the sorting process is adapted to handle several criteria for which the sorting complexity naturally increases.

5.3.3 Application example – between two datasets

Figure 5-3 and Figure 5-4 illustrate, respectively, how close (belonging to the same class) or disjointed (belonging to different classes) various sets can be at the Forsmark site. In this example, the results are presented twice: (1) on a classical linear diagram (left column); and (2) on a linear-log diagram (right column). The second representation highlights the relative importance of the standard error bars with regard to their simultaneous density values for small and large densities. This better reflects the role of each pole zone value in the distance calculation (Equ. 5-4), in which each term has the same weight, whatever the density.

Sets KFM08C-FFM01a and KFM08B-FFM01 belong to a single class (with $d_{max} = 1$) for each of the three observation grids. Their global densities (Equ. 4-6) are very close (Figure 5-3*a*), but, also, when refining the computation according to dip only (Equ. 4-7), or according to the *13-pole-zones grid*, the similarity remains. Note that by increasing the number of pole zones in the grid, we observe some local variations, such as that depicted in pole zone 6 in Figure 5-3*c*. In contrast, a class consistency for a refined grid probably will lead to consistency with regard to less refined grids. In any case, two similar datasets, which, according to the three grids, are very close to one another, can be defined as belonging to the same FD. In the present example, both sets already belong to FD FFM01. Among the 60 datasets defined for Forsmark with $d_{max} = 1$, only four couples of datasets are associated with single classes for the three grids.

Alternately, the sorting process leads to discrimination between datasets with significant differences, as is observed for KFM01A-FFM01 and KFM01A-FFM02 (Figure 5-4). In this case, almost all density values appear disjointed, with a clear, global density gap. The only way to group these two datasets in the sorting process would be to relax the threshold criterion (increase d_{max}), which is equivalent to a systematic increase in standard error bars associated with density values. Doing so could be worthwhile in the context of a site analysis, in order to identify close datasets with respect to others that are much different (even if, strictly speaking, the merging criterion is not matched).

In the following section, the sorting process is applied to site analyses. Both Forsmark and Laxemar sites are considered.

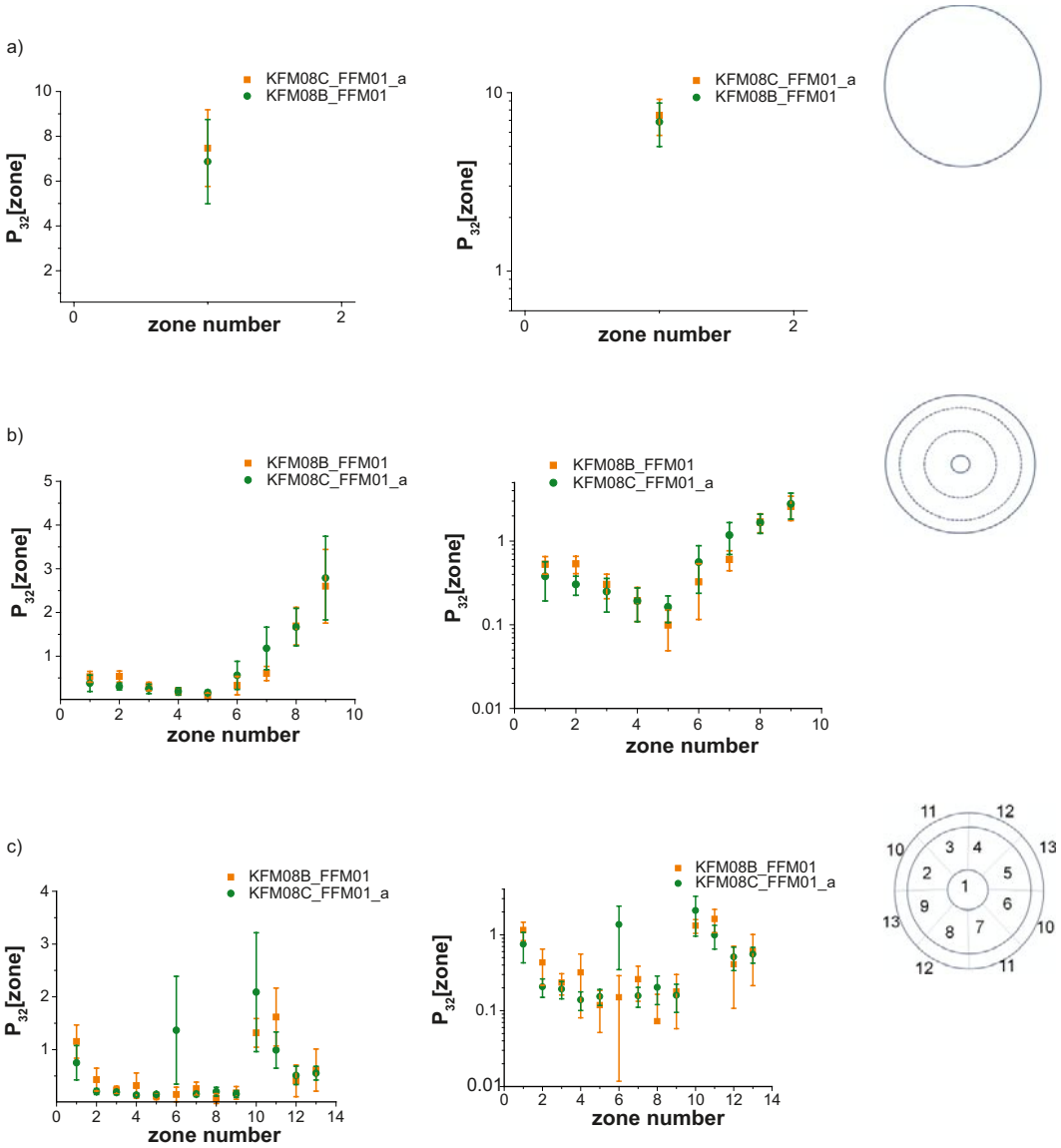


Figure 5-3. *Datasets KFM08C_FFM01_a and KFM08B_FFM01 belong to the same class for the three grids. The first column shows regular representation, and the middle column gives linear-log representation.*

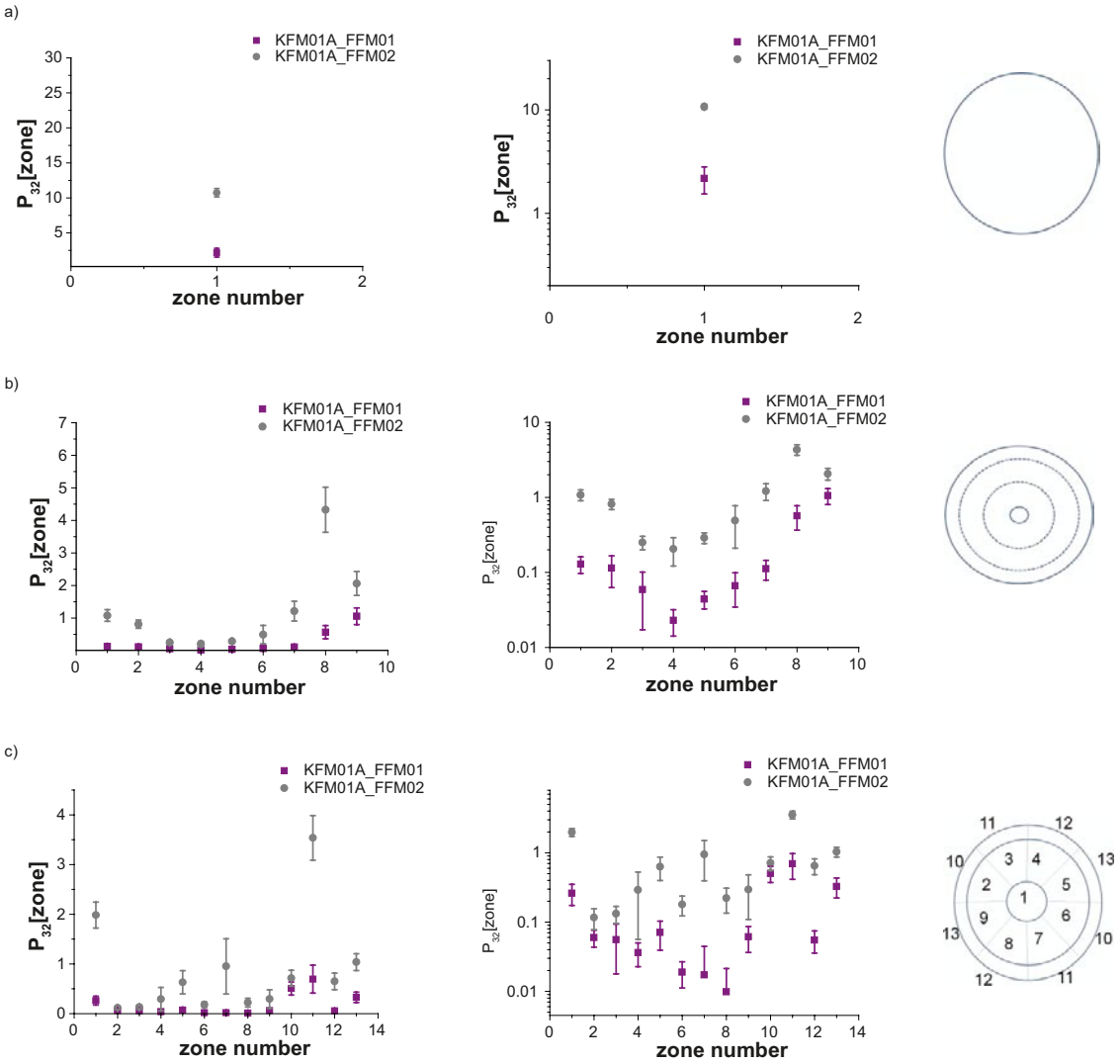


Figure 5-4. *Datasets KFM01A_FFM01 and KFM01A_FFM02 belong to a distinct class for the three grids. The first column shows regular representation, and the middle column gives linear-log representation.*

5.3.4 Control of the sorting process

The sorting algorithm is based on point to point density distributions comparison. If only one point defines the density (*1-pole-zone grid*), then the distance calculation fully reflects the matching between distributions. In complement, as soon as the number of points defining the distribution is larger than 1, we get a "global matching": locally, a quite large distance can be compensated for by another very close distance at another location of the distribution. The merging criterion is fulfilled if, on average, the distance, point-to-point, from two distributions, is smaller than the allowed threshold. Therefore, the *13-pole-zones grid* comparison allows more local discrepancies than would, obviously, the *unique-pole -zone grid*, which is based only on one point.

The class definition arising from the sorting process also is related partly to the selected grid of observation and to the level of constraint reflected through d_{max}. In practice, we can fix the final number of classes, or d_{max}, required for the classification. The resulting class robustness then is assessed further (Section 5.2.3.4). Besides, a user-defined initial classification can be imposed to take direct account of the FD division at the site scale, for example, and relevancy can be assessed next.

The evolution of d_{max} with the number of classes is displayed in Figure 5-5.

First, refining the grid of observation – at constant d_{max}, from the *1-pole-zone grid*, the *dip pole zone grid* and the *13-pole-zones grid*, results in increasing the number of classes. This reflects the heterogeneous nature of the statistical repartition of fracture pole orientations.

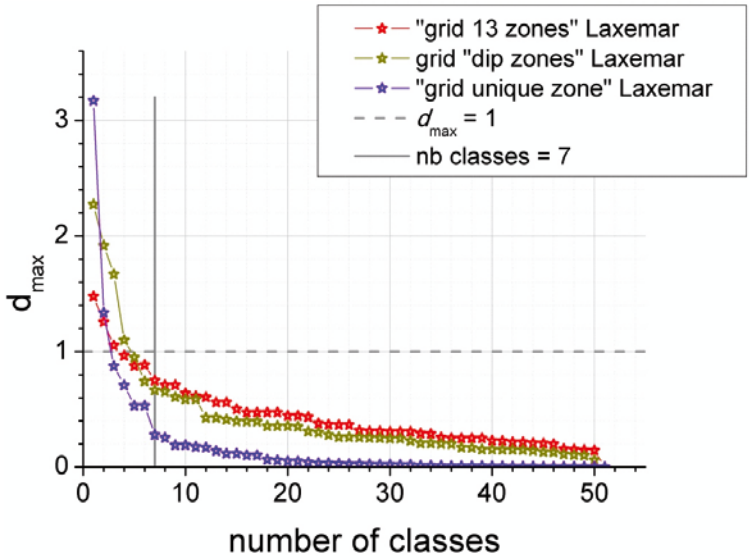


Figure 5-5. *d_{max} = f (number of classes) at the Laxemar site. The initial datasets are defined according to fracture domains and boreholes.*

In general, the number of classes starts (right to left in Figure 5-5) from the number of distributions, with $d_{max} = 0$, and goes up to more than a few units when finally all distributions are grouped into one big class:

- Having small d_{max} entails hardening the merging condition below the standard deviations allowed from the uncertainty and the variability analyses. Accordingly, for a very restrictive criterion, only two perfectly identical distributions can be merged. Because this case never occurs in practice, one gets as many classes as the number of initial distributions considered during the sorting process.
- From a certain point of view, allowing a larger d_{max} involves artificially increasing the variations allowed around mean values, which, in turn, allows to more distributions to be merged according to their relative closeness, even if their density distributions are not strictly compatible from a statistical perspective.
- Variations of d_{max} with the class number are rather smooth. This reflects the continuous nature of the fracturing throughout the site, rather than a clear discrimination into domains with very different statistical properties.
- For two adjacent values (n; $n + 1$), a significant shift of d_{max} values reinforces the relevancy of classification "n + 1 classes" with regard to the "n classes". This is illustrated in Figure 5-5 for the *13-poles-zones grid*, between 5–6 classes on one side and 7 on the other.
- Having $d_{max} = 1$ reflects that, **on average,** the distributions and average distribution into a class overlap through their standard deviations (Figure 5-2*a*). Having $d_{max} = 0.5$ reflects a higher overlapping with density values and a range defined by standard deviation values.

Therefore, the general trend logically combines an increase of d_{max} with a decrease of the final number of classes. In practice, an optimized classification level is chosen according to the occurrence of a drop of d_{max} values between the range 0.5 and 1.

The dominating continuum behavior mentioned above is illustrated further through **global parent densities,** expressed by P_{32}, and computed from the *1-pole-zone grid*. The sorting process between distributions relies on one single value per distribution. The analysis leads to a comparison of single density values. The density distributions and their sorted classifications are displayed in Figure 5-6. The **global parent densities**, expressed by P_{32}, range between 4 and 25 m^2/m^3. The obtained classification, which ends up with 7 classes for $d_{max} = 0.3$, is reflected through the different colors in the plot.

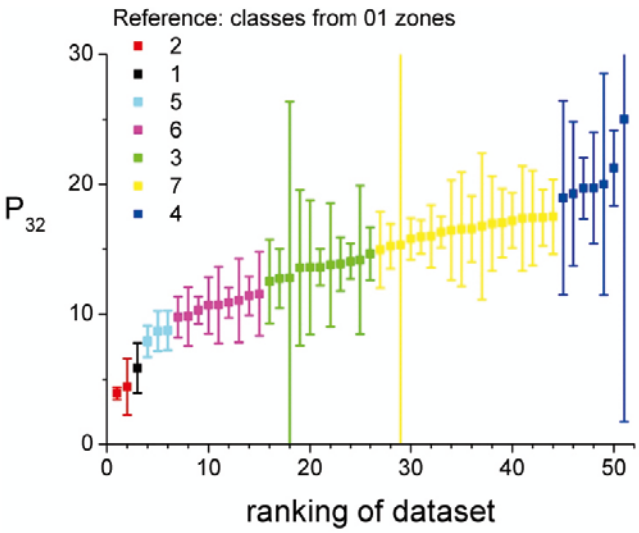


Figure 5-6. *Global parent density (integrated over all orientations) and associated standard deviations for each dataset at the Laxemar site. A rank (x-axis) is used to sort densities according to increasing values (in y-axis). Statistical fracture domains are represented by colors (see inset).*

The repartition is correlated perfectly with the mean density value (Figure 5-6) following an increase trend. However, the variation of density is smooth from one set to the other. In such cases, one cannot determine easily, or observe locally, why divisions occur at a specific location. In particular, the limit, in terms of density, between Class 3 and Class 7 is not governed by a local shift of density but, rather, by the global behavior of both classes. Whereas a clear gap between density values can be identified easily, continuous variations raise additional difficulties: in the present example, all datasets from Classes 3 and 7 cannot be merged with the given threshold of d_{max} set to 0.3, even if local variations are not striking. This observation underlines the difficulty of defining independent domains.

The second point illustrated in Figure 5-6 refers to the resulting standard deviation, σ_{P32} (error bars in the graph), associated with each density value. This figure displays a large spectrum – from very tightened around the density value itself and above. In the first case, densities are well defined; in the second case, as long as the standard deviation is equal to or larger than the mean density value itself, density cannot be defined. It happens twice over 51 datasets, for KLX09C_FSM_N and KLX11A-FSM_W, respectively, ranks 29 and 18 in Figure 5-6. This arises from conditions illustrated in Figure 4-4 and Section 4.3.5. It reflects the combination of very large uncertainties and large orientation bias.

5.3.5 Class formation from distributions

In this section, we show how the grouping of distributions is structured and how it evolves with d_{max}.

When a distribution is represented by more than one point (as in Figure 5-6) , it becomes difficult to display how grouping evolves given the sorting constraints. For instance, in Figure 5-7, several distributions, all belonging to one single class in the final classification associated with the Laxemar site, are plotted. It is difficult to understand the classification logic from diagrams that are too complex.

To create better **visualization**, we use a method from /Cox and Cox 1994/ to display a projection of all the studied distributions in the 2D space using multidimensional scaling (MDS). MDS is a widely used technique for visualizing data. It maps high-dimensional data in a low-dimensional representation, mainly 2D or 3D. The MATLAB Toolbox for Dimensionality Reduction has been used to compute the 2D coordinates /van der Maaten 2007/. The process is applied in Figure 5-8. (See Appendix 3 for additional details). We expect that distributions gathered into one class will appear clustered in the projection.

The above figure also highlights the two-fold nature of distributions in the sorting process. It shows how the final class repartition (Figure 5-8*b*) is composed of strong trends, with groups of distributions permanently clustered together (from small d_{max}, Figure 5-8*a*) and other distributions that are aggregated progressively, and whose links to one cluster or another may evolve.

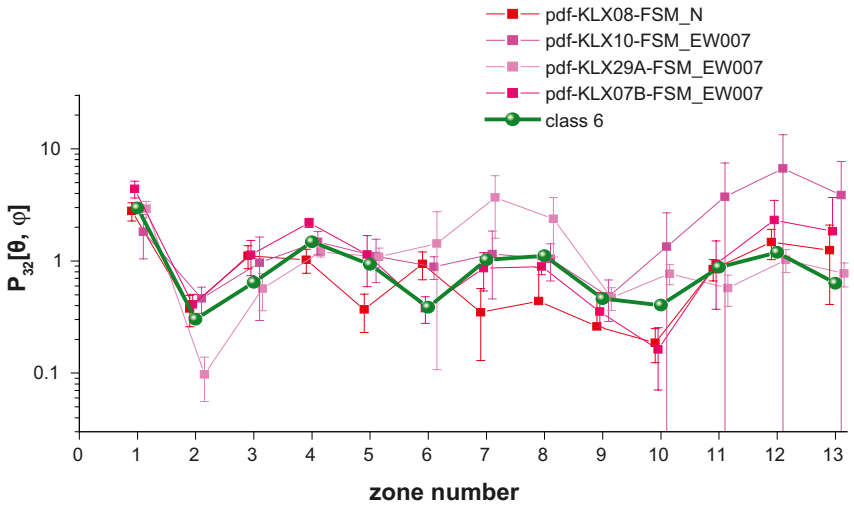


Figure 5-7. *Distributions belonging to Class 06 – 13-pole-zones grid (rose-to-red colors); mean distribution associated to Class 06 in green.*

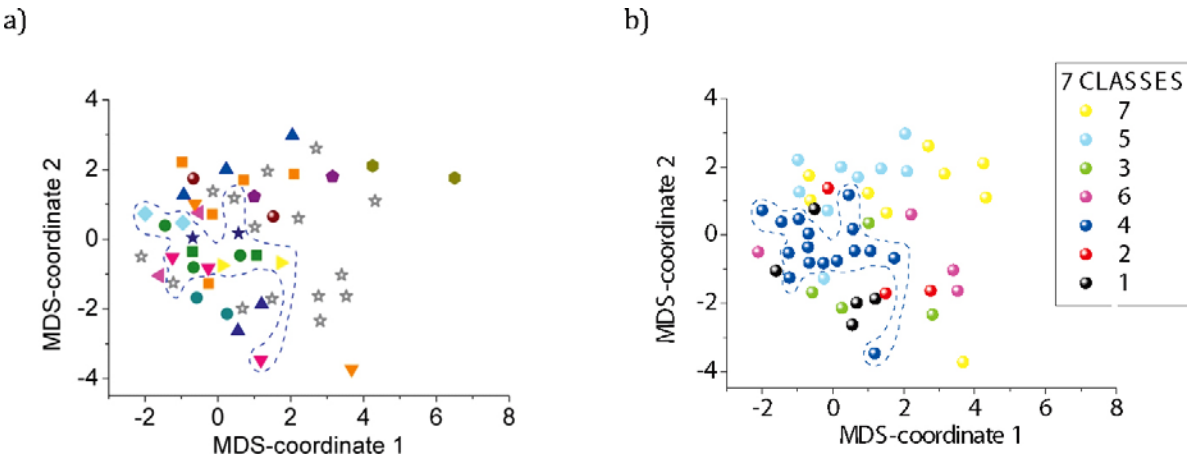


Figure 5-8. *Fifty-one (51) density distributions from the database at the Laxemar site computed from the 13-pole- zones grid and projected onto 2D-space according to the MDS method /Cox and Cox 1994/: **a)** The colored filled symbols reflect clusters of distributions identified with d_{max} = 0.4 and consistent with d_{max} = 0.8 (These can be viewed as permanent/strong clusters through the sorting process.) The open star symbols indicate distributions whose class membership varies with increasing d_{max}. **b)** Colored balls refer to one of the 7 classes from the final classification chosen.*

5.4 Application to the Laxemar site

5.4.1 Initial conditions

From the site model defined by SKB, six FDs are currently defined at Laxemar: FSM_N, FSM_S, FSM_C, FSM_W, FSM_EW007 and FSM_NE005. The borehole database is accordingly divided, by reference to both Fracture Domain and borehole belonging, leading to 51 initial datasets (Figure 3-2). Once done, most of the resulting datasets contain more than 100 discrete fracture poles.

Such an initial partition relies therefore on local fracturing and geological properties (single hole interpretation); however possible links between boreholes or closeness criteria are not considered. In complement, two initial conditions are next tested: essentially, the sorting process is initiated while ignoring the possible link between datasets coming from the same FD (5.4.2.1 to 5.4.2.3). Finally we consider the possibility to start the sorting process from a partition copied on the FDs (5.4.2.4).

We have shown in the previous sections how a final classification is potentially sensitive to the accuracy of the grid of analysis and to the level of allowed variation. In the present case, we chose to keep the grid with 13 pole zones and to consider a classification that results in 7 main classes with $d_{max} = 0.8$ (Figure 5-5). This leads to a number of statistical classes of the same order as that defined at Laxemar by SKB for the FD. We provide here initial results of the application of the fully quantitative method that has been developed in the framework of the current project.

5.4.2 Results

5.42.1 Site partition into statistical classes

Sorting the dataset distributions up to a threshold $d_{max} = 0.8$ leads to group them into 7 classes relatively to the *13-pole-zones grid*. The statistical robustness of each class is further assessed in next section 5.4.2.2. The thus formed site partition is summarized in Table 5-2 below.

The partition obtained differs from the FDs defined. Except for class 2, at least two to four different FDs occur in each statistical class. This means that statistical fracturing properties are not sufficient for identifying the FDs. This point is further investigated in section 5.4.2.4 and discussed in Chapter 6. The class density distributions and stereonet representations are summarized below in the two Figure 5-9 and Figure 5-10.

In the present partition, the number of datasets by statistical class varies from 3 to 16, whereas the class weight goes from 2–3% up to almost one third of the total available core length. Class 4, which contains the largest number of datasets, thus reflects close to 30% of the core length investigated. As displayed in Figure 5-9 (thick red line) and in Figure 5-10*a*, it encompasses higher densities for the preferential subvertical fractures and for close to horizontal fractures. Class 7 displays a very similar profile (magenta line in Figure 5-9 and Figure 5-10*b*), with horizontal and subvertical fractures at the same level and clearly above the middle zones (highlighted by a back light gray box in Figure 5-9).

Table 5-2. Laxemar site. Site partition into 7 classes for the case of the *13-pole- zones grid*, $d_{max} = 0.8$. The number of datasets by class is provided, as well as the corresponding dataset names, the cumulative core length portion and fracture number proportion.

Class ref number (Total P_{32})	Datasets by statistical class	Number of datasets	Cumulated core length (%)	Cumulative fracture number (%)
1 (16.3)	KLX07A-FSM_EW007, KLX12A-FSM_NE005, KLX07A-FSM_NE005, KLX10B-FSM_EW007, KLX10C-FSM_EW007	5	9%	11%
2 (13.8)	KLX11E-FSM_W, KLX17A-FSM_W, KLX25A-FSM_W	3	6%	7%
3 (11.0)	KLX05-FSM_NE005, KLX18A-FSM_C, KLX18A-FSM_EW007, KLX21B-FSM_NE005	4	16%	15%
4 (14.3)	KLX04-FSM_EW007, KLX09G-FSM_N, KLX15A-FSM_C, KLX24A-FSM_W, KLX04-FSM_N, KLX11A-FSM_W, KLX16A-FSM_S, KLX26A-FSM_NE005, KLX07B-FSM_N, KLX13A-FSM_W, KLX22A-FSM_W, KLX26B-FSM_NE005, KLX09-FSM_EW007, KLX14A-FSM_W, KLX22B-FSM_W, KLX28A-FSM_NE005	16	31%	36%
5 (11.0)	KLX09B-FSM_N, KLX09F-FSM_N, KLX20A-FSM_W, KLX09C-FSM_N, KLX11B-FSM_W, KLX09D-FSM_N, KLX11C-FSM_W, KLX09E-FSM_N, KLX11D-FSM_W	9	7%	8%
6 (12.4)	KLX07B-FSM_EW007, KLX08-FSM_N, KLX10-FSM_EW007, KLX29A-FSM_EW007	4	2%	3%
7 (6.3)	KLX03-FSM_C, KLX09-FSM_N, KLX23B-FSM_W, KLX04-FSM_C, KLX10-FSM_C, KLX23A-FSM_W, KLX08FSM_C, KLX11F-FSM_W, KLX08-FSM_EW, KLX19A-FSM_W	10	28%	20%

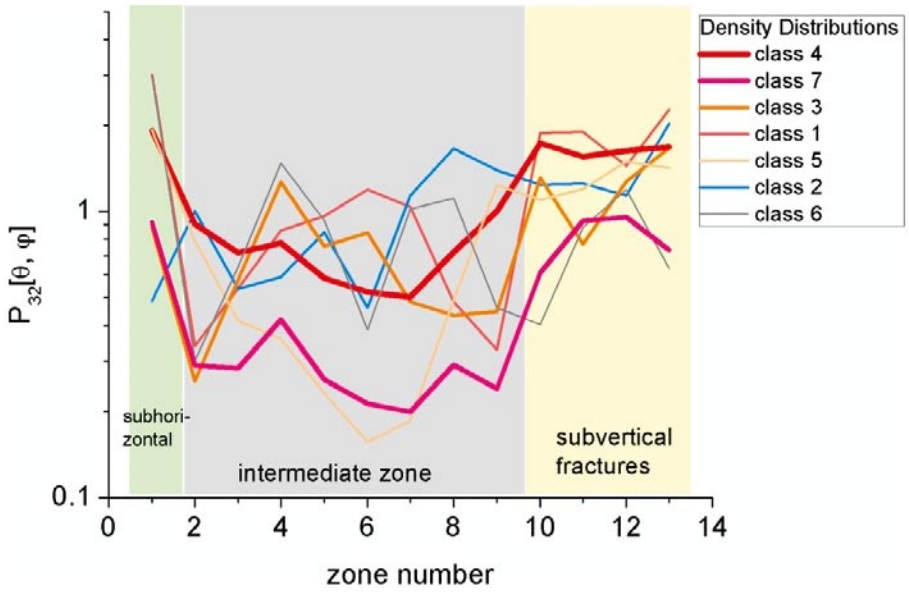


Figure 5-9. *Mean class distributions from Laxemar: 7 classes, d_{max} = 0.8, 13-pole-zones grid. Pole zone 1 refers to the central zone (subhorizontal fractures), pole zones 2 through 9 refer to the middle zones, and pole zones 10 to 13 refer to subvertical fractures (Figure 4-1).*

There is however a clear shift in total density (total P_{32} for class 4 is 14.3 whereas is it 6.3 for class 7) which prevents distributions from both classes to be merged into a single one. In contrast, the distinction between class 4 and class 1 is less obvious: indeed, both classes display close P_{32} values, similar subvertical fracturing and finally also a horizontal preferential orientation (Figure 5-10*a* and Figure 5-10*c*). However, for class 4 the subhorizontal are close to strike 15 whereas they are close to strike 180 in class 1. Due to this discrepancy, the densities calculated in the middle differ and lead to differentiate datasets from both classes. The remaining classes are more clearly different from the 3 abovementioned.

Class 5 displays the highest contrast between different pole zones, with a maximum density obtained for the subhorizontal pole zone. We note that this class mainly combines two particular sets of datasets: the KLX09B, -C, -D, -E and -F on the one side and the KLX11B, -C and -D. Both groups start just below the surface. Also, the last dataset of the statistical class corresponds to elevation between –50 and –150 m. Class 2 and class 3 display the same preferential subvertical EW direction, but they differ for second order peaks. Note that class 2 groups 3 datasets from a single FD, the FSM_W. Finally, class 6, which represents less than 5% of the database coverage , displays a strong preferential orientation in the WE direction.

Previous interpretations highlight that real differences exist between the several statistical classes. Note that for the interpretation process, the representation of both the density distribution curves (Figure 5-9) and the contoured stereonets (Figure 5-10) are highly complementary. Indeed the stereonets quickly allow to visually identify preferential directions and relative density variations within a single dataset, whereas the density distribution curves better express absolute density variations. The above interpretations also show that different classes share partial similarities and striking differences. The next section shows how to quantify this overlapping and therefore how to assess the statistical relevance of the classes.

We have performed here an initial analysis, for which the pole zone grid chosen is regular and independent of the site specificity. In view of the first round of results, we could consider updating the pole zone analysis grid in order to better catch the site specificity. However, one has to start with this systematic analysis. For instance, once the most systematical preferential directions on the fracture directions are identified the grid can be accordingly updated.

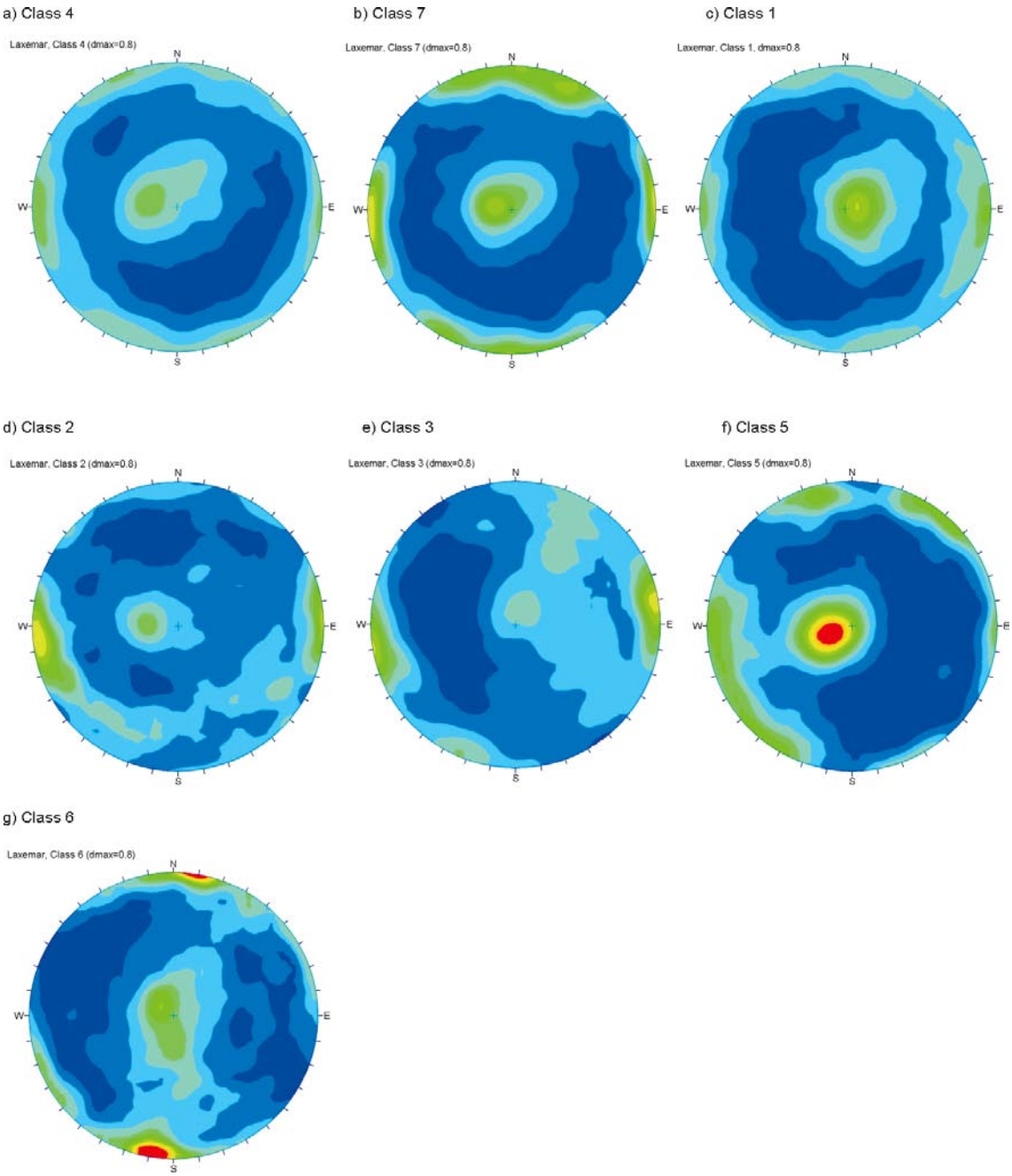


Figure 5-10. *Contoured stereonets (including orientation bias correction) for a) Class 4, b) Class 7, c) Class 1, d) Class 2, e) Class 3, f) Class 5, g) Class 6. The composition of each statistical class is recalled in Table 5-2. Once a statistical class is defined from several datasets, all the corresponding fracture poles are summed into the stereonet representation, including their respective orientation bias correction.*

5.4.2.2 Statistical robustness associated to classes

The coherence analysis aims at defining and quantifying to which extent each dataset falls into the best adapted statistical class and finally to which extent classes have a real existence. In doing so, coherence indexes are computed from Equ. 5-8 and Equ. 5-9. As defined in Equ. 5-7, the probability of each dataset distribution to belong to a given class is computed. Even if each distribution does belong to a particular class, it could have a non-null probability of belonging to another. Computations are based on distances from one class to another. In order to get a first overview of the coherence analysis, the probability of each distribution (y-axis) belonging to one of the classes (x-axis) is plotted in the form of a contour map (Figure 5-11). In the plot, high probabilities of occurrence are represented in orange-red, lower probabilities in blue. Also, the representation is such that distributions from one class are grouped (see the figure legend for details).

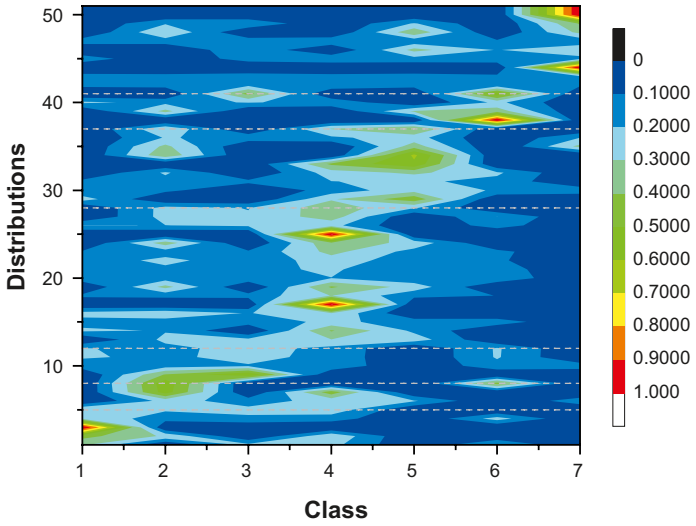


Figure 5-11. *Contoured plot of the probability for a distribution (y-axis) to belong to each class defined (x-axis). Distributions from 1 to 51 are sorted according to their class of reference, from 1 to 7. (Dashed grey lines indicate the limit between classes – i.e. distribution rows 1 through 7 refer to Class 1 and so on).*

Probabilities are computed after the sorting process: in other words, the sorting process initially tries to define the typical characteristics of fracturing throughout the site. This is done by the determination of the class mean distribution. Once a classification is established, these probabilities reflect how well the available data are represented by it. The two aspects, of course, are correlated, but this allows better assessment of the representativeness of the classification. We observe that, usually, the maximum probability associated with a distribution effectively falls into its own class. A full coherence analysis takes account all distributions and all classes, with and without weighting according to the number of distributions within each class (which varies from 3 to 16 for Classes 2 and 4, respectively).

Again, because the general properties of fracturing throughout the site cannot be easily sectioned into discrete domains but, rather, belong to a wide continuum (of continuously varying properties), grouping many distributions into a class also induces a decrease in the individual representativeness of each element of the class.

Finally, we quantify the **classification strength** by considering the coherence parameters defined in Equ. 5-8 and Equ. 5-9. Both parameters arise from a combination of the above-mentioned probabilities; they are plotted in Figure 5-12 and Figure 5-13.

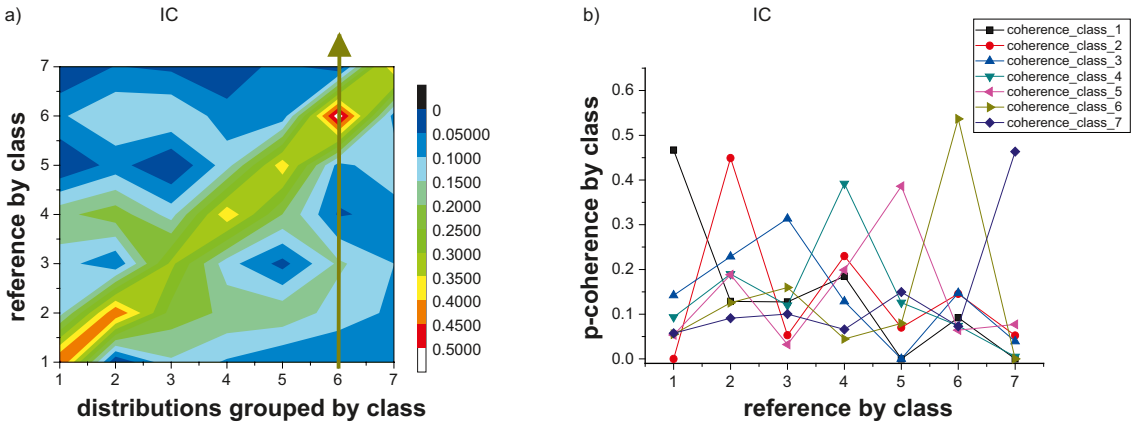


Figure 5-12. *Index of coherence IC by class: a) Contour map of the mean probability value; b) Classical x-y diagram representation. A curve is formed by the successive probability for a group of distributions (say, for instance, all distributions from Class 6, in brown) to belong to each class. A curve in **b**) refers to a vertical line in **a**), see, for example Class 6 symbolized by the brown arrow.*

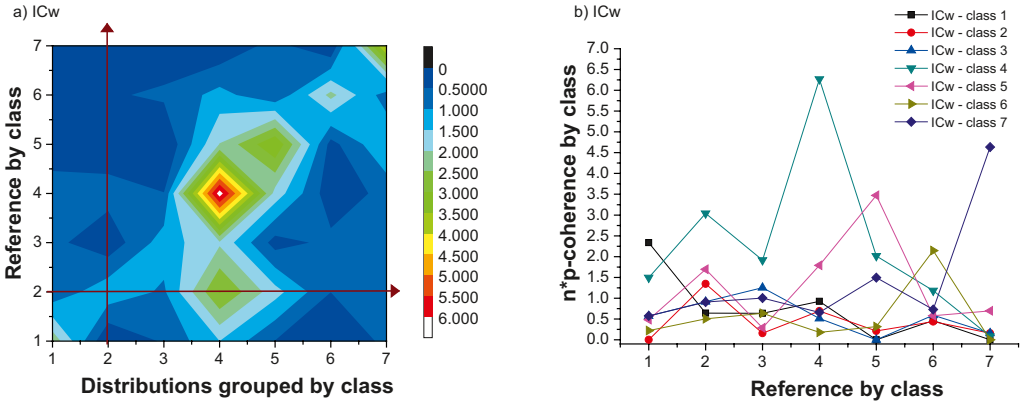


Figure 5-13. *Weighted index of coherence, IC_w, from Equ. 5-9: a) contour map; and* ***b****) classical x-y diagram. Plotted values refer to the total "number of distributions" remaining in each class when overlapping is taken into account.*

For all distributions in a class, we consider their mean probability (Equ. 5-8) to belong to one of the classes formed – i.e. their own class or another one. If the probability thus defined is significantly higher for its own class than for others, then the mean density distribution of the class can be viewed as a strong reference throughout the site. Accordingly, we observe from the contour map plotted in Figure 5-12*a*, higher values along the diagonal (where "reference by class" = "distributions grouped by class") and lower values in the remaining terms, which are representative of overlapping between classes. For a set of distributions linked to a class, one can observe the IC indices in Figure 5-12*b*, plotted as curves. Note that Figure 5-12*a* and *b* are two representations of the same data. The highest IC is obtained for Class 6, but all classes display the same trend.

The IC criterion expresses a degree of confidence in the class definition without encompassing the basic number of distributions involved in each class. Thus, IC neglects the initial total number of distributions encompassed into a class. Indeed, if a class is associated with a single distribution, its mean distribution will be exactly equal to the distribution, leading to a probability of 1 to belong to the class. In contrast, a class containing several distributions will inevitably encompass more variability, which presumably will decrease its internal coherence. Therefore, the second consistency index, IC_w, is weighted by the respective number of distributions into a class. This gives a direct level of representativeness for each class from the initial set of distributions and allows comparison of various classifications from an initial set.

Resulting values are plotted in Figure 5-13. As observed above, the stronger values still are observed for the diagonal on the contour map (Figure 5-13*a*).

In the chosen representation, the sum of the values along a vertical arrow in Figure 5-13*a* will result in the "real number of remaining distributions" in a class. (This is also the value obtained when summing the values by class in Figure 5-13*b*). Next, a sum along the horizontal (Figure 5-13*a*) will provide the initial number of distributions grouped in order to form the classes (i.e. directly arising from the sorting process). Both quantities are plotted in Figure 5-14. We observe a high increase in Class 2 and a relative decrease in Class 4; no significant change is noticed in the other classes, although some reorganization occurs according to overlapping. This reflects the high degree of overlapping between distributions and classes (which is a reality). Increasing the number of distributions into a class results in a relative decrease in class precision, leading to the relative weight decrease evaluated here.

5.4.2.3 Sensitivity to level of detail considered

In Figure 5-15, we illustrate how the classification arising from the previous *13-pole-zones grid* of analysis is reflected partially in the simpler unique pole zone analysis. Recall that the unique pole zone is based on a single value (equivalent to P_{32}) per dataset, reflecting the total density, independent of orientation. The unique pole zone analysis provides classes sorted according to increasing, global parent densities. The 13 pole zones classification is more complex than a simple additional subdivision of initial classes. Therefore, the distribution-distance criterion does not result in exactly the same division; the global parent density is no more discriminating when orientation is taken into account.

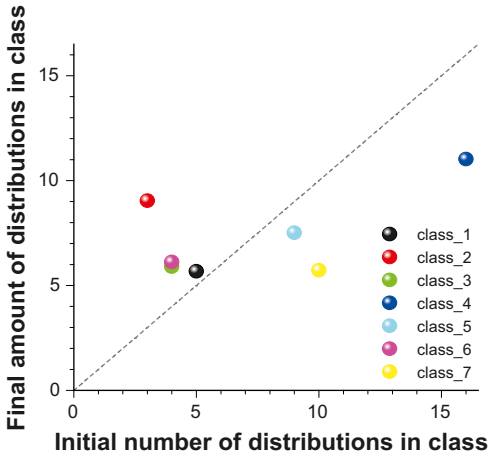


Figure 5-14. *Readjustment between the number of distributions associated with defining a class and the final "number" of distributions in the same class once overlapping has been included.*

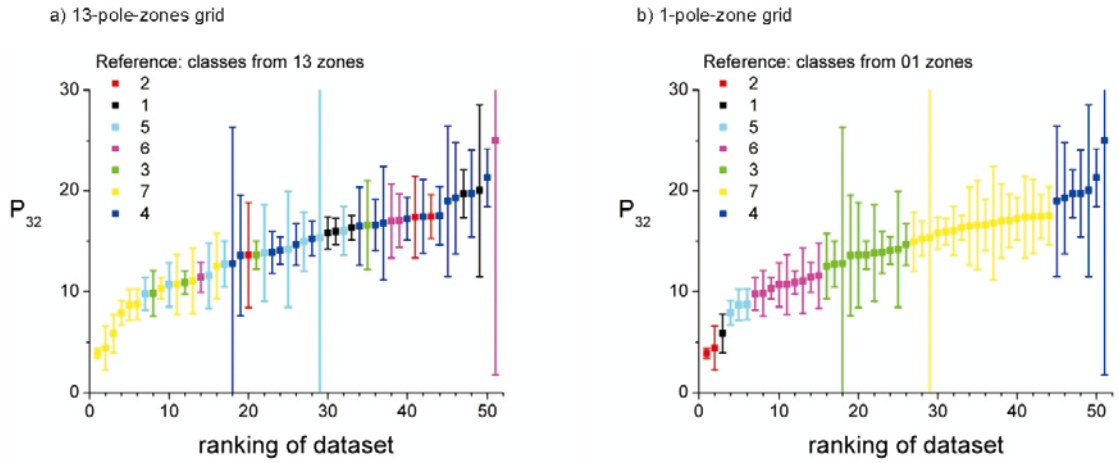


Figure 5-15. *Class identification from the global* $\mathbf{P}_{32}$ *representation:* ***a)*** *classification from the* ***13-pole-zones grid****;* ***b)*** *exact reproduction of Figure 5-6b.*

However, the behavior observed from *1-pole-zone* is respected, only roughly, for the 13 pole zones analysis. (On average, Class 7 has a global density lower than Class 5 etc up to Class 4 and 6). Again, this illustrates the high degree of overlapping between defined classes and the complexity of fracturing properties distribution throughout the site.

5.4.2.4 Comparison with statistical robustness of the Fracture Domains partition

The Fracture Domain Division is encompassed partly into the previous statistical analysis, as basic datasets are divided accordingly. However, the results of the analyses have shown large discrepancies between the final distribution repartition into classes and the Fracture Domain Division: all defined classes gather distributions from different FDs (Table 5-2).

Here we simply force the sorting process to end up with a classification equivalent to the Fracture Domain Division and perform the corresponding coherence analysis. Results are provided in Figure 5-16. The Forced Fracture Domain Classification gives $d_{max} = 1.51$ (larger than $d_{max} = 0.8$ obtained above). We observe that the Fracture Domain Division does not result in coherence indexes that are high enough. In fact, the coherence indexes attributed to classes (i.e. the diagonal in Figure 5-16*a*) are not dominant, except the one associated to class FSM_S; however, this class contains only one distribution, leading to a coherence index equal to 1.

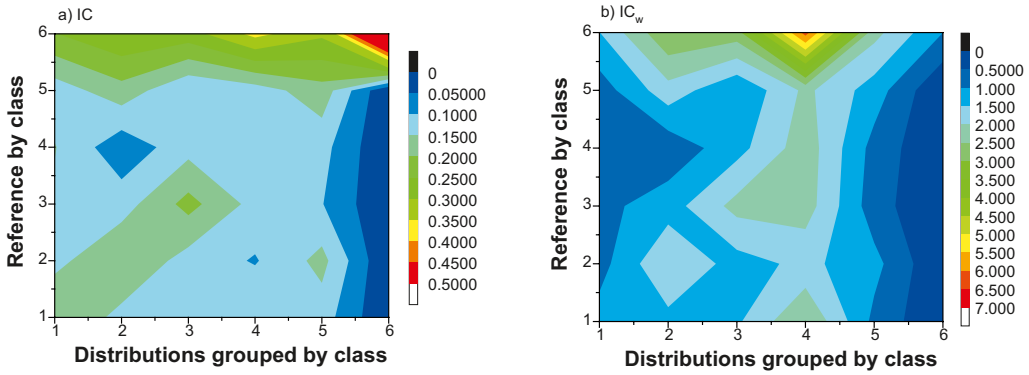


Figure 5-16. *Index of coherence IC (Equ. 5-8) and weighted index of coherence IC_w (Equ. 5-9) for a classification based on the FD reference. With class 1 for FSM_C, class 2 for FSM_EW007, class 3 for FSM_N, class 4 for FSM_W, class 5 for FSL_E005 and class 6 for FSM_S.*

Therefore, the Fracture Domain Division probably is not based on a systematic and quantitative comparison of each part of the parent density/orientation distributions. Some aspects of the density/orientation probably have been highlighted, whereas others have been neglected.

The method developed here is quantitative and encompasses all fractures from a database, including quantification of orientation bias, uncertainty and variability. As a counterpart, until now we have considered only the full range of orientations in an analysis, giving equivalent weight to each pole of a dataset. In other words, two distributions are matched if all their points are close. This is perhaps the source of the discrepancies observed between the FD and the statistical sorting. Moreover, a spatial criterion has been used in creating the FD, and this information has not been included in the statistical analysis.

5.5 Application to the Forsmark site

Due to time constraints, this section has not been developed to the same extent as has the section on the Laxemar site.

From the site model defined by SKB, ten FDs are characterized at Forsmark: FFM01, FFM01_affected_by_DZ, FFM02, FFM02_affected_by_DZ, FFM03, FFM04, FFM04_affected_by_DZ, FFM05, FFM05_affected_by_DZ and FFM06. Accordingly, the borehole database is divided by reference to both fracture domain and borehole, leading to 60 initial datasets (Figure 3-2). Most of the resulting datasets contain more than 100 discrete fracture poles.

The evolution of d_{max} with the number of classes for the three types of grids considered is displayed in Figure 5-17.

We consider a final classification that results in 11 classes (see Appendix 4); the corresponding coherence indexes, IC in a) and ICw in b), are displayed in Figure 5-18. Clear domination of diagonal values is observed, thus validating the proposed classification. As was noted from the Laxemar site analysis, the statistical classes in Forsmark gather distributions from different FDs.

Finally, we force the classification to fit the major Fracture Domains (six classes from FFM01 to FFM06). Two cases are considered, one with inclusion of the “affected by DZ” with their corresponding “normal” group (leading to a total of 6 FDs) and another where they are differentiated (leading to 10 FDs). The coherence analysis is illustrated in Figure 5-19.

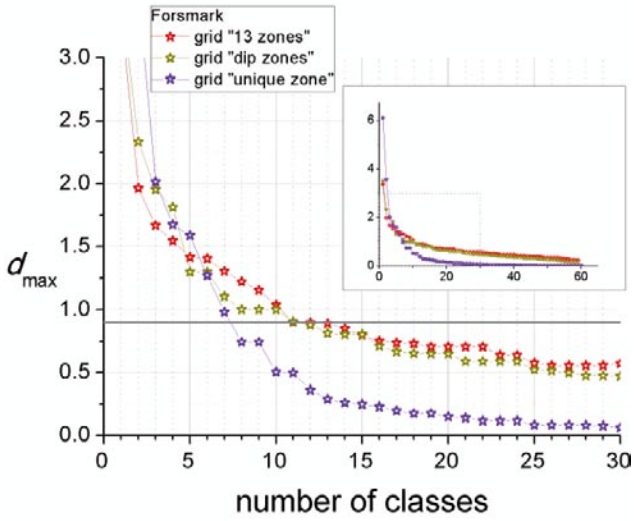


Figure 5-17. *d_{max} = f (class) at the Forsmark site. Datasets defined according to fracture domains. The inset provides the full range of variation and the main plot focus on smaller d_{max}.*

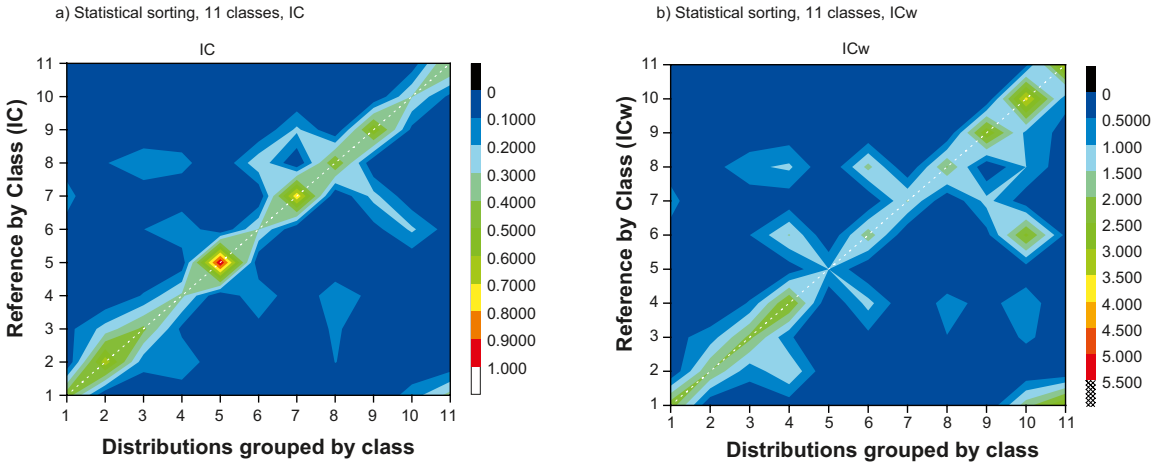


Figure 5-18. *Coherence indexes from the final Forsmark statistical sorting (d_{max} = 0.9, 11 classes).*

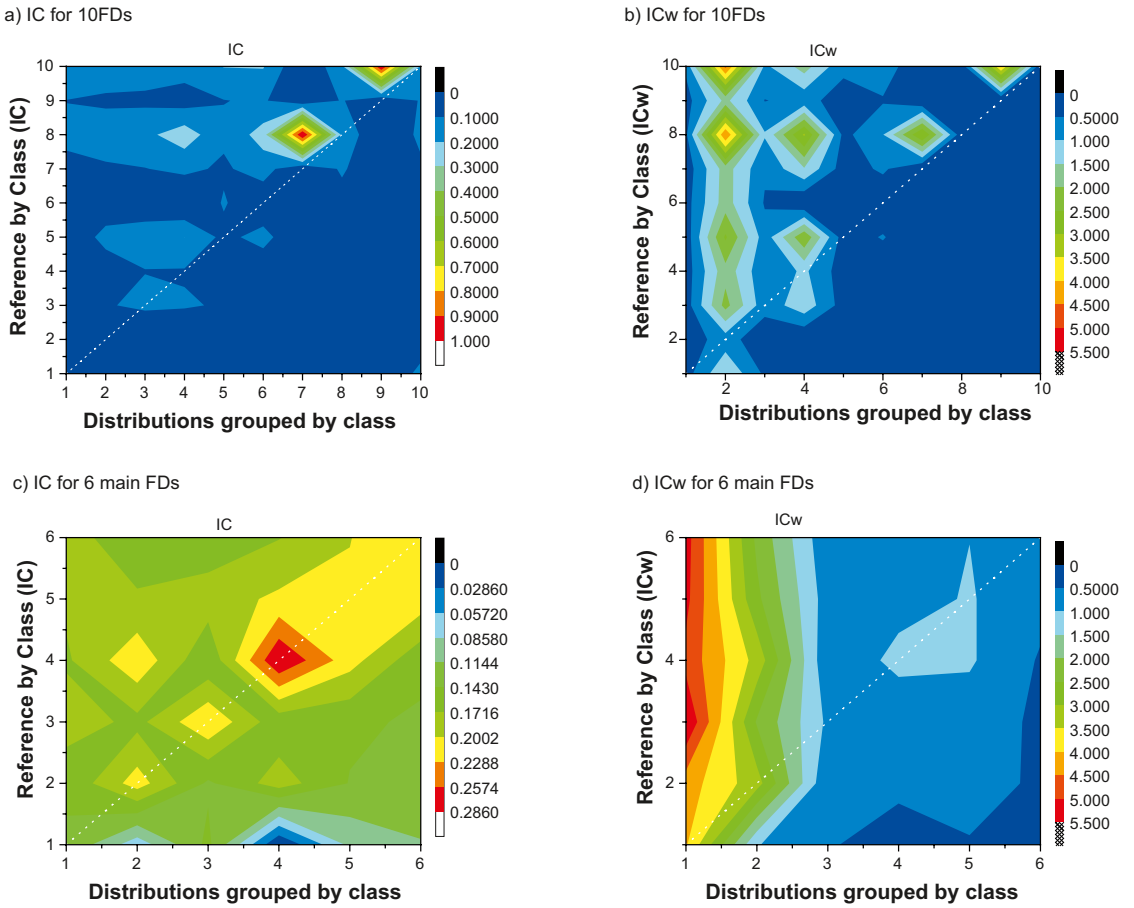


Figure 5-19. *Index of coherence, IC (Equ. 5-8) and weighted index of coherence, IC_w (Equ. 5-9) for a classification based on the fracture domain reference at Forsmark: a) and b) for ten fracture domains; c) and d) for six main FFM domains.*

5.6 Comparison between the Laxemar and Forsmark sites

Despite the difficulty mentioned above in defining a single classification per site, we can compare general trends at the two sites. We first recall that densities generally are larger at Laxemar than at Forsmark (Figure 5-20 and Figure 4-9). In both cases, the density evolution from one dataset to the another is rather smooth. Therefore, modeling tasks at both sites are confronted with the difficulty of defining clear boundaries between Fracture Domains. This effect gets more pronounced when the level of refinement in the comparison (i.e. density + orientations) is increased.

In general, for a given d_{max} , the Laxemar site tends to gather more datasets than does the Forsmark site, resulting in less classes (Figure 5-21). This can be interpreted as the Laxemar site having higher homogeneity. However, several boreholes at the Laxemar site were drilled very close to one another, which can be the reason for the apparent higher homogeneity.

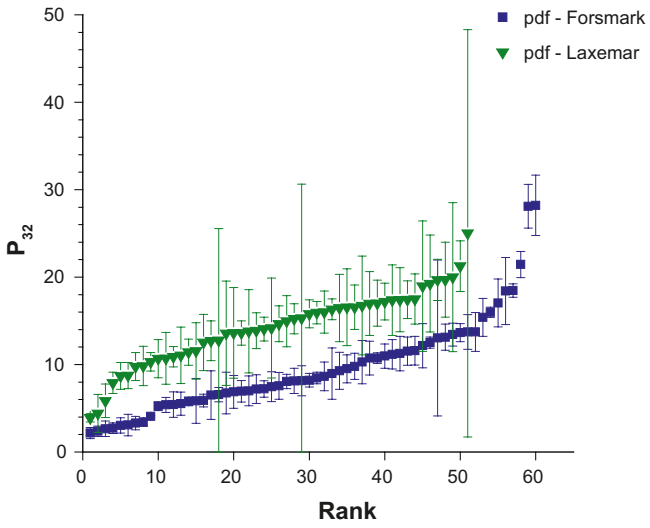


Figure 5-20. *Global parent densities at both sites, expressed by* P_{32}. *Values are plotted according to their rank with regard to increasing mean density values.*

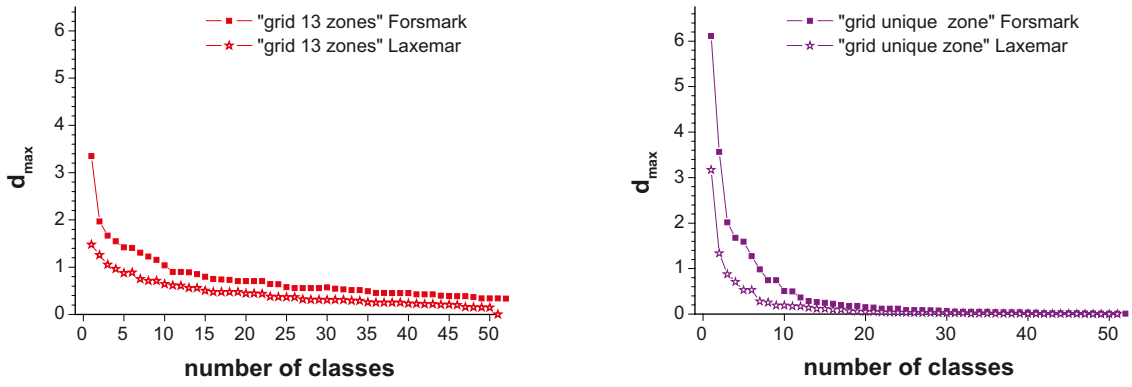


Figure 5-21. *Comparative evolution of* $d_{max} = f$ *(number of classes) between both site databases:* ***a)*** *13-pole-zones grid,* ***b)*** *1-pole-zone grid.*

5.7 Conclusions from the statistical fracturing properties analyses

Depending on the fracturing level of complexity, the final objective of the analysis was to derive a certain number of well-identified *DFN models* (sets of parameters) in order to statistically describe fracture properties throughout a site. Accordingly, our basic idea was to allow a **systematic** and **quantitative** comparison between various datasets, encompassing uncertainties arising from limited knowledge and restricted access to data. The comparison between datasets thus must account for the relative **uncertainties** and **variability** attached to the data. For that purpose, we have developed and applied a complete methodology.

One critical aspect of the proposed method is to define a relevant quantity for describing fracturing and to make it comparable from one dataset to the other, beyond the recurrent observation biases in this field of application. To this point, we have focused on the **parent density/orientation distributions** that can be computed from the numerous borehole data. This was described in Section 4. In short, the parent density associated with any arbitrarily defined range of pole orientations can be computed. In practice, a regular grid is chosen. A grid element can be defined, for instance, by all poles with dips lower than 30° (referring to subhorizontal fracturing). In addition, given the uncertainty of its original orientation, each fracture pole contribution to the density distribution can spread over more than 1 grid pole zone.

An automated sorting process is then built. Any number of parent density/orientation distributions is taken as input. This process results in a classification in which initial datasets are gathered according to their level of compatibility, thus forming classes from which mean properties are derived. The distance between distributions, therefore, is defined. Two distributions are merged if, on average, their point-to-point density distribution discrepancies differ less than a given threshold (d_{max}). At that point, preliminary analyses highlight the heterogeneous nature of the pole strike repartition. As a consequence, a rather refined grid (of 13 pole zones) has been chosen to perform additional analyses. In such cases, the distance between two classes reflects the mean distance associated with each grid pole zones thus defined. Also, the distance computed incorporates an inverse weighting of the density by its related standard deviation, this in order to reflect the loss of knowledge due to large uncertainties. Finally, the resulting density distributions associated with classes can be viewed as the main trends for describing the site properties. Initial density distributions are combined to compute the class properties; reversely, at the end of the process, the probability for each initial distribution to belong to any of the final classes thus formed reflects the level of overlapping (or inversely, the strength of each class to represent the general trends over the site investigated).

When applied to the available site databases, the first and most striking outcome of the sorting process is reflected in the **high level of overlapping between classes**. In other words, no clear gap (from the perspective of statistical fracturing) appears between datasets recorded at various site locations. The statistical properties of fracturing throughout the site should not be seen as a combination of domains with clear boundaries, but, rather, as a large continuum of permanently varying properties. Nevertheless, it remains that several classes or main trends can be defined. Observation of the formation along the sorting process (while increasing d_{max}) highlights two main distribution behaviors: (1) some distribution clusters are defined very clearly from rather small values of the threshold; and (2) other distributions act as "satellites" sometimes attracted by a well-defined cluster. Beyond the variability and uncertainties encompassed in the modeling process, according to the level of detail required at a given stage, a significant portion of the distributions can be attracted by one class or another, which is reflected through the class coherence indices computed elsewhere.

On the method we conclude that:

- In general, both coherence indices defined in this study are complementary: the first gives the intrinsic coherence of a class; and the second defines the proportion of distributions finally combined in a class. Note that the latter quantity still provides a limited view of the site: indeed, the various distributions considered here, linked originally to various datasets and underlying samples, do not reflect either the initial sample sizes or their spatial repartition.
- The domain definition, or sorting procedure, involves grouping various locations having different statistical properties. Each domain is viewed as a type of pole within which merging is acceptable, rather than being viewed as underlying identical fracturing properties.
- The proposed classification relies on a statistical criterion and is consistent. At this stage, however, it cannot be defined as the unique and best division.

The final classification at the Laxemar site contains 7 classes identified on the *13-pole- zones grid* with $d_{max} = 0.8$. In these cases, coherence indices are significant, thus validating the relevancy of the classification proposed. Increasing d_{max} would have led to fewer classes (Figure 5-5) with higher intrinsic variability, thus decreasing the internal relevancy and IC (coherence indices). We emphasize that if the proposed classification is relevant from a statistical point of view, it does not guarantee spatial continuity between various domains and deserves refinement. Moreover, although based on the same local definition of core sections and datasets, the statistical classes we obtain differ from the Fracture Domain model provided by SKB. In the next chapter – Conclusions- we open the discussion and propose future directions of work in order to bridge the gap between strict statistics and complexity of a site.

6 Conclusions and perspectives

In this work, we have tried to better constrain the DFN modeling framework for a site-scale application. The main issues investigated have concerned the scaling nature of fracture size distribution, the relative weight of properties assessed with regard to uncertainty and variability, and division of the site into consistent *statistical classes of similar fracturing properties*. In the report, detailed conclusions have been inserted at the end of each chapter (2.6 Physical rationale of a Universal model of Fracture Scaling, 4.7 Conclusions on the variability and uncertainty and 5.7 Conclusions from the statistical fracturing properties analys). We briefly recall the main outcomes:

- The scaling properties of the fracture size distributions are in good agreement with an underlying physical model in which a fracture can grow freely to the point that another fracture inhibits its growth. The model predicts a self-similar behavior above a certain scale (probably close to a few meters), whatever the scaling exponent and density at smaller scales. Connectivity and hydraulic properties of such particular systems are not known.
- In most cases for the Laxemar and the Forsmark sites, variability is predominant over uncertainty. In addition, neither variabtility nor uncertainty can completely mask information. It is possible, therefore, to determine significant values for DFN model parameters.
- A fully quantitative method has been developed to sort fracturing properties from any number of datasets according to their statistical similarity. Thus, site complexity can be seen at one glance. Potentially, the method is highly flexible. It has been used in this work to compare, systematically, all orientations from all datasets. It also can be used to focus on more specific aspects (see below).

From the sorting process outcome interpretations, we observe:

- A high level of overlapping between classes or defined *statistical classes of fracturing*. The statistical properties of fracturing throughout the site should not be seen as a combination of domains with clear boundaries but, rather, as a large continuum of permanently varying properties. The challenge is to differentiate the strong or “attracting” fracturing properties at some locations from the rest.
- A clear discrepancy between our statistical classes and the existing FD model. This is a sign that different criteria were considered to form the classes. We perform a purely statistical comparison based only on quantitative fracturing parameters. In our understanding, the current FD model arises from both statistical and non-statistical constraints. This results in a large intrinsic variability within each FD. How can a methodology reconcile this more geologically oriented approach with as rigorous a statistical treatment as possible? Note that as far as the geological model is uniquely defined, the statistical methodology may be applied within each "geological" unit.

Since various DFN models are expected from the various Fracture Domains, the statistical relevancy of their definitions is important. A classification as defined in Chapter 5 is an objective comparison of fracturing properties, governed by the minimization of intrinsic variability within a class. It does not arise from a clustering of continuous fracture domains. In order to adapt it further to the DFN modeling specificities of the SKB sites and in general to the conditions inherent to geological models, we propose several tracks: first of all, the initial *13 zones grid* used may not be the most appropriate. For instance at Laxemar, we notice that *middle pole zones* (between horizontal and vertical poles) are rarely dominant and rather are filled due to their proximity with horizontal or vertical fracturing. But they represent more than half of the points used for computing the distances between density distributions. In complement to the grid shape optimization, we would consider the possibility to introduce a weight proportional to the density in the pole zone. Secondly, we would revisit the initial division into FDs and DZs especially in order to better consider the role of DZs, which are currently masked in the analysis (they are straightaway removed from analyses). Finally, since the similarity and discrepancy between various density distributions may be partial (two datasets may share the same density of horizontal fracturing whereas they differ for other directions), it can be worth considering independently restricted ranges of the fracture datasets (according to orientations).

Nonetheless, we recommend to initiate the sorting process and consecutive analyses by a regular grid which allows an objective observation of the fracturing properties.

7 Index and Notations

Some abbreviations and expressions are recalled here:

DFN	**Discrete Fracture Network.**
FD – Fracture Domain	– SKB nomenclature – In /Olofsson et al. 2007/: Fracture Domain (FD). A fracture domain refers to a rock volume outside deformation zones in which rock units show similar fracture frequency characteristics. Fracture domains at Forsmark are defined on the basis of the single-hole interpretation work and its modifications and extensions, and the results of the initial statistical treatment of fractures. Both deformation zones that have been modelled geometrically, and possible deformation zones that have been identified in the single hole interpretation but have not been modeled geometrically, are excluded from fracture domains. The term is used in the first hand as a basis for the discrete fracture network modelling work (geological DFN) and different fracture domains at Forsmark are referred to as FFMxxx. The recognition of fracture domains as defined here aims primarily to meet the needs of collegues working with the disciplines hydrogeology, hydrogeochemistry and rock mechanics.
DZ – Deformation Zone	– SKB nomenclature – In /Olofsson et al. 2007/: Deformation zone (DZ). A deformation zone is a general term referring to an essentially 2D structure along which there is a concentration of brittle, ductile or combined brittle and ductile deformation. The term fracture zone is used to denote a brittle deformation zone without any specification whether there has or has not been a shear sense of movement along the zone. A fracture zone that shows a shear sense of movement is referred to as a fault zone. In the single-hole interpretation, deformation zones are referred to as DZxx, where the ID is coupled to the boreholes.
Local Dataset or local Fracture Domain	At depth the initial division of boreholes into subsets or local datasets relies on the local Fracture Domain, or the former single hole interpretation, definition. For instance, KLX18A-FSM_C or KFM06C-FFM05. Once DZ have been removed, the local datasets rely on (pseudo)-continuous core sections.
SFD or Statistical Class	A Statistical Fracture Domain is made of a group of datasets whose density distributions can be represented by a unique DFN model set of parameters. It is not strictly equivalent to FDs.
Pole zone or angular zone	Binning corresponding to a range of fracture pole orientation.
13-pole-zones grid	Based on the binning of the orientation distribution into 13 regular pole zones, see Figure 4-1*c*.
PDF	Probability Density Function.

In addition, the following notation definitions are recalled:

Table 7-1. Main notations.

	Fracture distribution model
$\mathbf{N_{3d}(\theta,\varphi,l,L)}$	Reference name for the Fracture Density Distribution model. (Equ. 4-1).
	Bracket **(**θ,φ**;** ***l*****)** indicates dependency on strike θ, dip φ**,** size ***l*** and system size *L*.
$\boldsymbol{n_{3d}(l,\theta,\varphi)}$	Reference name for the Normalized Fracture Density Distribution model. (Equ. 4-2).
	Bracket **(**θ,φ**;** ***l*****)** indicates dependency on strike θ, dip φ and size ***l***.
$\boldsymbol{n_{1d}(\theta,\varphi)}$	One- dimensional (1D) fracture distribution model. (Equ. 4-3).
Subscript *3d*	Refers to model parameters expressed in "full space", ie in 3 dimensions or Euclidian dimension *d* equal to 3.
α_{3d}	Density term of the fracture distribution model. (Equ. 4-3).
$\alpha_{3d}(\theta,\varphi)$	Density term specifically referring to a range of orientations, defined in brackets: (θ,φ) for strike/dip range, (θ) for a strike range and (φ) for a dip range.
$\alpha_{3d}(\theta)$	
$\alpha_{3d}(\varphi)$	
σ_{unc}	Standard deviation associated to density term α_{3d} and arising from uncertainties associated to fracture intercept orientations. Subscript "unc" refers to uncertainty.
σ_{var}	Standard deviation associated to density term α_{3d} and arising from the variability analysis of a dataset. Subscript "var" refers to variability.
σ_{3d}	General/generic term used to represent the standard deviation associated to density term α_{3d}, when the thus defined density distribution is the input of the sorting process. Used in Chapter 5.
$\boldsymbol{r_o}$	Location parameter.
$\boldsymbol{r_{min}}$	Minimum fracture radius consistent with the DFN model.
$\boldsymbol{l_{min}}$	Minimum fracture diameter, $l_{min} = 2*r_{min}$.
l_c	Arbitrary cutoff diameter in a DFN model ($l_c \geq l_{min}$).
L	Typical size of a system (e.g. side *L* of a cube).
a	Exponent of a power-law size distribution, also called "length/size exponent". ($k_t = a_{2d} - 1$; $k_r = a_{3d} - 1$).
k	Shape parameter associated to the probability density function of a power-law distribution.
$\boldsymbol{k_r}$	Shape parameter associated to the probability density function of a power-law distribution of fracture radii.
$\boldsymbol{k_t}$	Shape parameter associated to the probability density function of a power-law distribution of fracture traces.
D	Fractal dimension associated to a DFN.
D_{1d}, D_{2d}, D_{3d}	Same as *D*, subscript refers to a dimension precisely measured resp. over a one-dimensional, two-dimensional or three-dimensional support.
Ψ (*â*)	Orientation bias correction (Equ. 4-4).
P ($\mathbf{k_r}$,*â*)	correction term related to orientation bias (Equ. 4-3).
p, $\mathbf{p_\varphi}$	Generic notation to referring to probability (Equ. 4-8, Equ. 4-9).
C, F, J	Notation used to represent an event in the probabilistic development.
	Fracture geometrical parameters.
Subscript i	Refers to a fracture intercept i.
r	Radius of a disc-shape fracture.
l	Fracture diameter for a disc-shaped fracture ($l = 2*r$).
t	Fracture trace size.
β	BETA (field from table p_frac_core_eshi).
σ_β	UNCERT_BETA (field from table p_frac_core_eshi).
â	ALPHA (field from table p_frac_core_eshi) (not noted α in order to avoid confusion with the density term).
$\sigma_{\hat{a}}$	UNCERT_ALPHA (field from table p_frac_core_eshi).
θ	STRIKE (field from table p_frac_core_eshi).
φ	DIP (field from table p_frac_core_eshi).
σ_θ	Uncertainty on strike (see Equ. 4-13).
σ_φ	Uncertainty on dip (see Equ. 4-13).
	Borehole orientation parameters.
B	BEARING (field from table p_frac_core_eshi).
σ_B	BEARING_UNCERT (field from table p_frac_core_eshi).
I	INCLINATION (field from table p_frac_core_eshi).
σ_I	INCLINATION_ UNCERT (field from table p_frac_core_eshi).

Table 7-2. Miscellaneous.

2D	two-dimensional (fractures appear as traces *'t'*).
2d	Subscript notation for 2D.
3D	Three-dimensional (disc-shape fractures of radius '*r*' or diameter '*l*').
3d	Subscript notation for 3D.
P_{32}	Fracture density expressed like total fracture surface by unit of volume.
P_{10}	Fracture intensity expressed like the total number of fracture intersections by unit of length.
h	Size of a borehole section (Equ. 4-3).
d	Borehole diameter (Equ. 4-3).
s	Used in general to refer to the size of the portion of a borehole section (Equ. 4-14).

Table 7-3. Notations in Chapter 5, Table 5-1.

$\alpha_{3d}(\theta,\varphi)\pm\sigma_{3d}(\theta,\varphi)$	**Fracture distribution model arising from local datasets, elementary piece or object of the sorting algorithm.**
E	Set of objects.
n	Number of object in a set.
C	class collection (collection of non-void subset of E).
X, Y, Z	Elements of C.
x, y	Objects (or distributions) of class X, resp. Y.
M	Class collection.
M_0	Collection containing all singletons (objects) of E.
$d(X,Y)$	Distance between two classes. Represents the degree of heterogeneity of $Z = X \cup Y$.
i	iteration of the sorting process.
d_{max}	Threshold on the distance between classes d (X,Y).
Ncs	Number of distributions in a class *CS*.
$(\alpha_{3d};\sigma)_j^i$	i: reference to a distribution. j: reference to a element in the distribution.
α_{CS}, σ_{CS}	Equivalent density distribution associated to class *CS*. (Equ. 5-1, Equ. 5-2).
d (m,n)	Distance between two distributions *m* and *n*. (Equ. 5-4).
N_Z	Number of pole zone in a grid (Equ. 5-4).
d_{cs}	Internal distance associated to a class (Equ. 5-5).
$d^*(x,yi)$	distance between a distribution *x* and a Class Y_i denoted by its mean distribution y_i (Equ. 5-7).
IC (X on X) = IC(X)	Internal Coherence index of class X (Equ. 5-8).
IC (X on Y)	Coherence index of class X in class Y (Equ. 5-8).
ICw	Weighted coherence index. (Equ. 5-9).

8 References

Ackermann R V, Schlische R W, 1997. "Anticlustering of small normal faults around larger faults." Geology 25 (12): 1127–1130.

Atkinson B K, 1989. Chapter 8. Fracture Mechanics of Rocks, Academic Press, Incorporated.

Bonnet E, Bour O et al, 2001. «Scaling of fracture systems in geological media.» Reviews of Geophyics 39 (3): 347–383.

Chelcea S, Bertrand P et al, 2006. A New Agglomerative 2–3 Hierarchical Clustering Algorithm Innovations in Classification, Data Science, and Information Systems. Berlin Heidelberg, Springer.

Cheng D, Kannan R et al, 2006. "A divide-and-merge methodology for clustering." ACM transcations on database systems 31 (4): 1499–1525.

Cox T, Cox M, 1994. Multidimensional scaling. London, UK, Chapman & Hall.

Darcel C, Davy P, Bour O, De Dreuzy J-R, 2006. Discrete fracture network for the Forsmark site. SKB R-06-79, Svensk Kärnbränslehantering AB.

Darcel C, Davy P, Bour O, De Dreuzy J-R, 2006. Alternative DFN model based on initial site investigations at Simpevarp. SKB R-04-76, Svensk Kärnbränslehantering AB.

Datta S, Datta S, 2006. "Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes." BMC bioinformatics 7 (397).

Davy P, Darcel C et al, 2006. "A note on the angular correction applied to fracture intensity profiles along drill core." J. Geophys. Res. 111: B11408.

Dershowitz W, Lee G et al, 1991a. FracMan Version 2.3. Interactive discrete feature data analysis, geometric modeling, and exploration simulation. User documentation. Redmond Washington USA, Golder Associates Inc.

Dershowitz W, Wallmann P et al, 1991b. Discrete fracture modelling for the Stripa site characterization and validation drift inflow predictions. Stripa Project Technical Report 91-16, Svensk Kärnbränslehantering AB.

Garai G, Chaudhuri B, 2004. "A novel genetic algorithm for automatic clustering." Pattern Recognition Letters 25 (2): 173–187.

Gordon A D, 1999. Classification, Chapman and Hall.

Hammah R, Curran J, 1998. "Fuzzy Cluster algorithm for the automatic identification of joint sets." Int J Rock Mech Min Sci 35: 889–905.

Hofrichter J, Winkler G, 2006. "Statistical analysis for the hydrogeological evaluation of the fracture networks in hard rocks." Environmental Geology 9 (6): 821–827.

Kim K, Yamashita E Y, 2007. "Using a K-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii". Journal of Advanced Transportation 41 (1): 69–89.

Le Goc R, Darcel C et al, 2008. "An automatic comparison method of probability density functions with uncertainties." In prep.

Mansfield C, Cartwright J, 2001. Fault growth by linkage: observations and implications from analogue models, Journal of Structural Geology, 23(5), 745–763.

Munier R, 1995. Studies of geological structures at Äspö. Comprehensive summary of results. SKB HRL Progress Report 25-95-21, Svensk Kärnbränslehantering AB.

Munier R, 2004. Statistical analysis of fracture data, adapted for modelling Discrete Fracture Networks-Version 2. SKB R-04-66, Svensk Kärnbränslehantering AB.

Munier R, Stigsson M, 2007. Implementation of uncertainties in borehole geometries and geological orientation data in Sicada. SKB R-07-19, Svensk Kärnbränslehantering AB.

Newman M, 2005. ”Power laws, Pareto distributions and Zipf’s law.” Contemporary Physics 46: 323–351.

Oltmans H, Verheijen P J T, 1997. “Influence of noise on power-law scaling functions and an algorithm for dimension estimations.” Physical Review E 56 (1).

Olofsson I, Simeonov A et al, 2007. Site descriptive modelling Forsmark, stage 2.2. A fracture domain concept as a basis for the statistical modelling of fractures and minor deformation zones, and interdisciplinary coordination. SKB R-07-15, Svensk Kärnbränslehantering AB.

Pickering G, Bull J M et al, 1995. ”Sampling power-law distributions.” Tectonophysics 248: 1–20.

Press W H, Teukolsky S A, Vetterling W T, Flannery B P, 2007. Modeling of data: Robust Estimation, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press.

Priest S D, Hudson J A, 1981. “Estimation of discontinuity spacings and trace length using scanline surveys.” Int. J. of Rock Mechanics and Mining Sciences 18: 183–197.

Sano O, Ito I, Terada M, 1981. Influence of Strain Rate on Dilatancy and Strength of Oshima Granite Under Uniaxial Compression, Journal of Geophysical Research, VOL. 86, No. B10, pages. 9299–9311.

Sarbu C, Zehl K et al, 2007. “Fuzzy divisive hierarchical clustering of soil data using Gustafson-Kessel algorithm.” Chemometrics and Intelligent Laboratory Systems 86 (1): 121–129.

Sornette D, Davy P, 1991. Fault growth model and the universal fault length distribution, Geophysical Research Letters, 18(6), 10.1029/91GL01054, 1079–1081.

Turcotte D L, Rundle J B, 2002. Self-organized complexity in the physical, biological, and social sciences, Proc. Natl. Acad. Sci. U. S. A., 99 (Suppl 1), 2463–2465.

van der Maaten L J P, 2007. An Introduction to Dimensionality Reduction Using Matlab. Maastricht, The Netherlands, Maastricht University. Technical Report MICC 07-07.

Vrac M, Stein M et al, 2007. “Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing.” Climate Research 34 (3): 169–184.

Appendix 1

Data resources for the fracture size analyses

We provide below two reference tables relative to fracture size analyses, Table A1-1 for the Laxemar site and Table A1-2 for the Forsmark site. For both sites, the lineament maps are further illustrated in Figure A1-1.

Table A1-1. Laxemar data resources for the size analyses.

Use	Apparent in mxd	Source	Comments
y	Outcrop fracture mapping ASM000025	SDEADM.GOL_SM_GEO_4121_VIEW	Outcrop trace maps
	Outcrop fracture mapping ASM000026	SDEADM.GOL_SM_GEO_4122_VIEW	
	Outcrop fracture mapping ASM000205	SDEADM.GOL_SM_GEO_4123_VIEW	
	Outcrop fracture mapping ASM000206	SDEADM.GOL_SM_GEO_4124_VIEW	
	Outcrop fracture mapping ASM000208	SDEADM.GOL_SM_GEO_4125_VIEW	
	Outcrop fracture mapping ASM000209	SDEADM.GOL_SM_GEO_4126_VIEW	
	Outcrop fracture mapping ASM100234	SDEADM.GOL_SM_GEO_4224_VIEW	
	Outcrop fracture mapping ASM100235	SDEADM.GOL_SM_GEO_4641_VIEW	
n	MDZ lineaments	SDEADM.SWP_LX_GEO_5583	
	MDZ lineaments (all southwest)	SDEADM.SWP_LX_GEO_5582	
y	Co-ordinated	SDEADM.GV_LX_GEO_4185	“Sub-regional lineaments”
y	Linked lineaments	SDEADM.GSF_SM_GEO_3282	“Regional lineaments”
y	Surface map	SDEADM.GOL_LX_GEO_6081	Modelled DZ

Table A1-2. Forsmark data resources for the size analyses.

Use	Apparent in mxd	Source	Comments
y	Outcrop fracture mapping AFM000053	SDEADM.GOL_FM_GEO_4781_VIEW	
	Outcrop fracture mapping AFM000054	SDEADM.GOL_FM_GEO_4084_VIEW	
	Outcrop fracture mapping AFM001097	SDEADM.GOL_FM_GEO_4087_VIEW	
	Outcrop fracture mapping AFM001098	SDEADM.GOL_FM_GEO_4086_VIEW	
	Outcrop fracture mapping AFM100201	SDEADM.GOL_FM_GEO_4801_VIEW	
y	Linked lineaments	SDEADM.GV_FM_GEO_6046	“Sub-regional lineaments”
y	Linked lineaments from co-ordinated lineaments	SDEADM.GV_FM_GEO_2079	“Regional lineaments”, version 1.2
y	Surface map	SDEADM.PSL_FM_GEO_5032	Modelled DZ

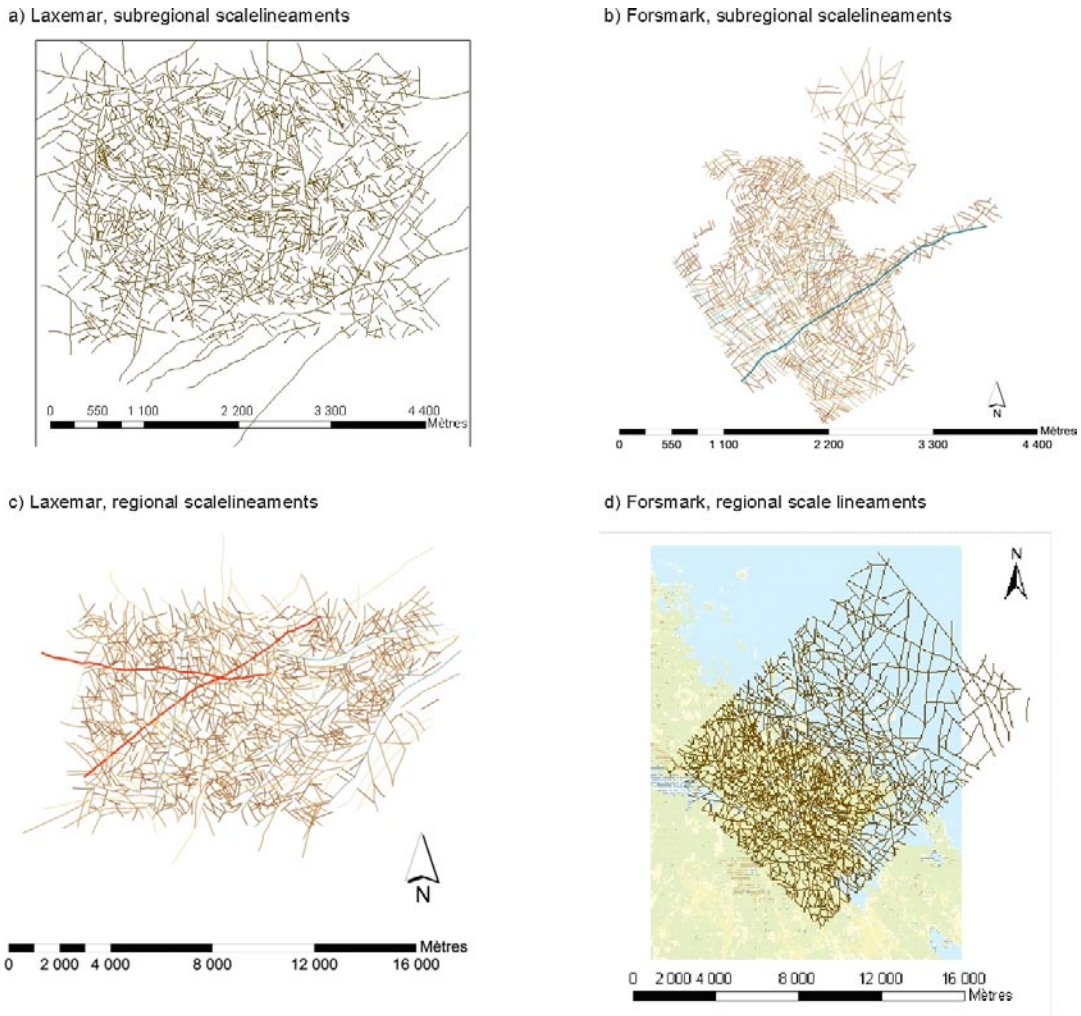


Figure A1-1. *Lineament maps. (a) and (c) at Laxemar site, and (b) and (d) at Forsmark site. "sub-regional lineaments" in (a) and (b), with maps of side close to 4,500 m and "regional lineaments" in (c) and (d), with maps of side close to 16 km. Note that in (c) large lineaments, in red, have a size between 9 km and about 11 km and in (b) largest lineaments, in blue, have a size between 3.5 and 4.2 km.*

Appendix 2

Data pre-processing

P_frac_core_eshi and p_freq_1m_eshi come from delivery BD_08_95.

In the uncertainty-variability analysis applied to the Forsmark site, all fracture intercepts (p_frac_core_eshi) having an activity GEO41 are pre-selected to be used as input in the following density distribution computation.

If the strike or dip column is null, the intercept is removed from the selection.

Appendix 3

Laxemar datasets, MDS visualization

The 51 datasets that were formed from the borehole data at Laxemar are shown in Table A3-1. In order to clarify the MDS visualization, a reference number is defined for each dataset. This number then is added to symbols that refer to the MDS projection onto MDS diagrams (Figure A3-1 and A3-2) . The evolution of the classification is illustrated further through MDS representation of the classification evolution with the d_{max} criterion. The MDS method allows simple projection in 2D space of the initial 13-point density-orientation distributions. The final projection (distribution positions in the diagram) depends only on the initial input dataset. We simply change the symbol colors to reflect the classification evolution.

Table A3-1. Dataset reference names and number – Laxemar site.

Distribution	File	Distribution	File
1	KLX03-FSM_C	31	KLX12A-FSM_NE005
2	KLX04-FSM_C	32	KLX13A-FSM_W
3	KLX04-FSM_EW007	33	KLX14A-FSM_W
4	KLX04-FSM_N	34	KLX15A-FSM_C
5	KLX05-FSM_NE005	35	KLX16A-FSM_S
6	KLX07A-FSM_EW007	36	KLX17A-FSM_W
7	KLX07A-FSM_NE005	37	KLX18A-FSM_C
8	KLX07B-FSM_EW007	38	KLX18A-FSM_EW007
9	KLX07B-FSM_N	39	KLX19A-FSM_W
10	KLX08-FSM_C	40	KLX20A-FSM_W
11	KLX08-FSM_EW007	41	KLX21B-FSM_NE005
12	KLX08-FSM_N	42	KLX22A-FSM_W
13	KLX09B-FSM_N	43	KLX22B-FSM_W
14	KLX09C-FSM_N	44	KLX23A-FSM_W
15	KLX09D-FSM_N	45	KLX23B-FSM_W
16	KLX09E-FSM_N	46	KLX24A-FSM_W
17	KLX09F-FSM_N	47	KLX25A-FSM_W
18	KLX09-FSM_EW007	48	KLX26A-FSM_NE005
19	KLX09-FSM_N	49	KLX26B-FSM_NE005
20	KLX09G-FSM_N	50	KLX28A-FSM_NE005
21	KLX10B-FSM_EW007	51	KLX29A-FSM_EW007
22	KLX10C-FSM_EW007		
23	KLX10-FSM_C		
24	KLX10-FSM_EW007		
25	KLX11A-FSM_W		
26	KLX11B-FSM_W		
27	KLX11C-FSM_W		
28	KLX11D-FSM_W		
29	KLX11E-FSM_W		
30	KLX11F-FSM_W		

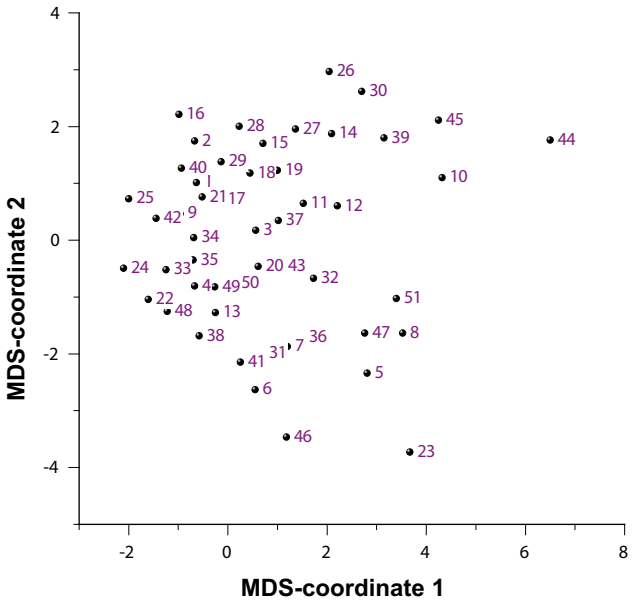


Figure A3-1. *MDS projection of the 51 density distributions of the Laxemar database.*

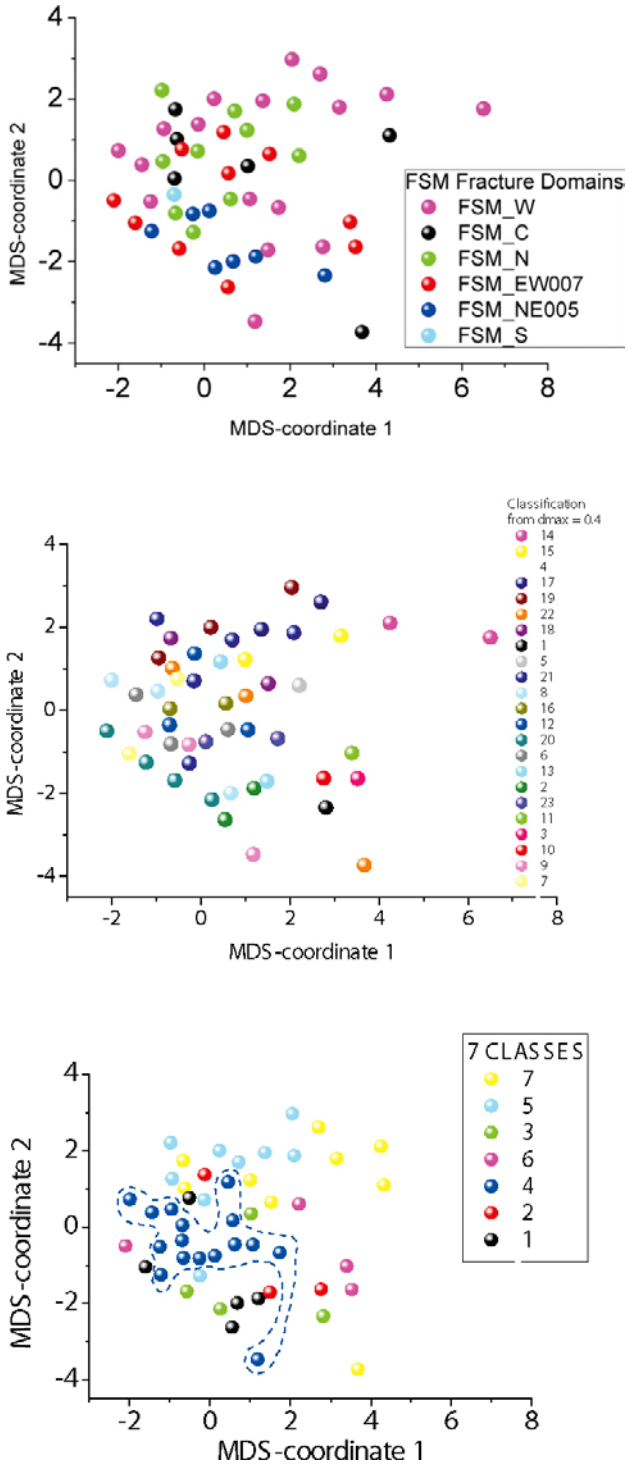


Figure A3-2. *Representation with MDS projection of the 51 distributions defined at Laxemar: (Top) color according to initial fracture domain; (middle) color according to sorting with d_{max} = 0.4; (below) color according to sorting with d_{max} = 0.8.*

Appendix 4

Forsmark proposed classification

Table A4-1. Forsmark classification into 11 classes.

Class reference number	Dataset names
1	KFM01B-FFM02 KFM01D-FFM02 KFM02A-FFM01_affected_by_DZ KFM07A-FFM01_a KFM09A-FFM01_a
2	KFM01C-FFM01_affected_by_DZ KFM09A-FFM04_affected_by_DZ KFM09B-FFM02_affected_by_DZ
3	KFM01A-FFM02 KFM01C-FFM02_a KFM01C-FFM02_affected_by_DZ KFM01C-FFM02_b KFM07C-FFM02
4	KFM04A-FFM04 KFM05A-FFM02 KFM06C-FFM05 KFM06C-FFM06 KFM07A-FFM05_affected_by_DZ KFM07B-FFM02 KFM10A-FFM01 KFM10A-FFM03
5	KFM08A-FFM01_c
6	KFM05A-FFM01_affected_by_DZ KFM06C-FFM01 KFM07C-FFM01 KFM09A-FFM01_b KFM09A-FFM04_a KFM09A-FFM05
7	KFM07A-FFM05 KFM09A-FFM04_b
8	KFM06B-FFM02 KFM07A-FFM02 KFM08A-FFM01_affected_by_DZ KFM09A-FFM02_affected_by_DZ
9	KFM06A-FFM06 KFM08A-FFM01_b KFM08A-FFM05 KFM08C-FFM01_b KFM08C-FFM06
10	KFM02A-FFM03 KFM03A-FFM03 KFM03B-FFM03 KFM04A-FFM01 KFM06A-FFM01 KFM06A-FFM02 KFM07A-FFM01_affected_by_DZ KFM08A-FFM01_a KFM08B-FFM01 KFM08B-FFM02 KFM08C-FFM01_a KFM09B-FFM01_affected_by_DZ
11	KFM01A-FFM01 KFM01B-FFM01 KFM01D-FFM01 KFM02A-FFM01 KFM05A-FFM01_a KFM05A-FFM01_b KFM07A-FFM01_b KFM07B-FFM01 KFM09B-FFM01

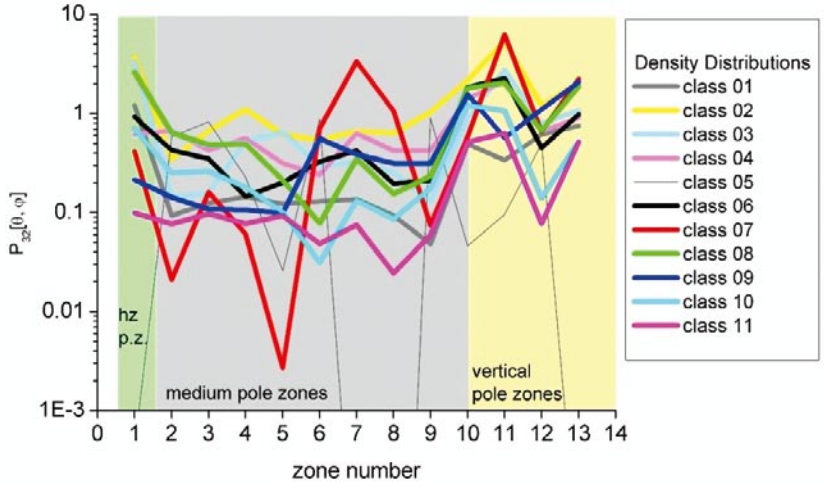


Figure A4-1. *Forsmark site. Class density distributions defined from the 13-pole-zones grid.*