Development of the statistical fracture domain methodology – application to the Forsmark site

Caroline Darcel, Romain Le Goc
Itasca Consultants SAS

Philippe Davy, Géosciences Rennes

December 2013

SKB

Development of the statistical fracture domain methodology – application to the Forsmark site

Caroline Darcel, Romain Le Goc
Itasca Consultants SAS

Philippe Davy, Géosciences Rennes

December 2013

Executive Summary

Fracture networks arise from the combination between fundamental fracturing processes and in situ conditions like geological factors, stress history, major faults, depth/pressure or rock mass properties. Changes of these conditions throughout space or time likely induce spatial variations of the fracture network geometrical properties. But the fracturing process itself induces consequent spatial variability which is not easily distinguished from the former. Thus, the complex spatial variability of fracture properties is a key issue for the DFN modeling.

This project aims to review and upgrade the site DFN modeling approach initially introduced in Darcel et al. (2009), especially with regards to relations between models, model assessment from data and spatial variability of properties. In particular, we aim to identify this difference in the observed variability: intrinsic to the fracturing process or extrinsic, that is, constrained by in situ conditions changes.

In this perspective the first stage of the work is dedicated to DFN models and modeling methodology. The commonly named Poisson DFN models are considered and their spatial variability fully defined. These models are then used as reference cases when real data are considered.

In the second stage real data from Forsmark site are analyzed to initially appraise their intrinsic density spatial variability and scaling behavior. This shows a typical signature of the real fracturing process which clearly differs from Poisson models and is closer to otherwise called fractal DFN models. Then the classification approach and the class properties assessment are applied to the Forsmark site and typical Statistical Fracture Models (SFM) are derived. The method is adapted to make objective analyses from any number of initial data. In the present case, close to 150 datasets with more than 60,000 data, are simultaneously considered into the classification process. The final classification identifies 5 to 6 major SFMs, whose definition strongly rely on the total density and to a depth dependence of the density of horizontal fractures.

Sammanfattning

Spricknätverk uppstår genom en kombination av fundamentala sprickprocesser och in situ förhållande såsom geologiska faktorer, spänningshistorik, större förkastningar, djup/tryck eller egenskaper hos bergmassan. Förändringar i dessa förhållanden över tid och rum kommer sannolikt att orsaka rumslig variation av spricknätverkets geometriska egenskaper. Dock orsakar sprick-bildningen i sig själv en rumslig variabilitet vilken ej med lätthet kan skiljas från den förra. Således, den komplexa rumsliga variabiliteten av sprickornas egenskaper är en nyckelfråga för diskret spricknätverksmodellering (DFN-modellering)

Detta projekt syftar till att gå igenom och uppdatera metodiken för DFN-modellering av plats-undersökningar som först presenterades i Darcel et al. (2009), i synnerhet när det gäller förhållandet mellan modeller, bedömning av modeller baserat på befintliga data samt den rumsliga variation av egenskaper. Framförallt vill vi identifiera denna skillnad i observerad variabilitet, den inneboende på grund av sprickbildningsprocessen eller den yttre, dvs begränsad av förändringar i in situ för-hållande.

Ur denna synvinkel redovisas och analyseras DFN modeller och modelleringsmetodik i första delen av arbetet. De så kallade Poissonska DFN-modellerna undersöks och deras rumsliga variabilitet är fullt ut bestämd. Dessa modeller används sedan som referensfall när verkliga data beaktas.

I den andra delen av arbetet analyseras verkliga data från Formarksområdet för att inledningsvis bedöma sprickdensitetens rumsliga variabilitet och skalningseffekt. Resultatet avspeglar en typisk signatur av den verkliga sprickbildningsprocess vilken tydligt skiljer sig från Poissonska modeller och istället mer liknar de så kallade fraktala DFN-modellerna. Därefter tillämpas klassificerings-strategin och klassegenskapsbedömningen på Forsmarkdata och typiska statistiska sprickdomäner, Statistical Fracture Models (SFM), utvecklas. Metoden är anpassad till att göra objektiva analyser från valfritt antal initialdata. I det föreliggande fallet beaktas närmare 150 dataset med mer än 60 000 data samtidigt i klassificeringsprocessen. Den slutgiltiga klassificeringen identifierar 5–6 statistiska sprickdomäner, SFM:er, vars gränser i huvudsak styrs av den totala sprickdensiteten och av djupberoendet av sprickdensiteten av horisontella sprickor.

Contents

1   Introduction

This project aims to further assess the robustness of the site DFN modeling approach initially presented in Darcel et al. (2009) and to particularly deepen the relation between possible Discrete Fracture Network (DFN) models and spatial variability of fracture properties. The method objectively sorts an initial amount of fracture datasets into an optimized set of "best" models. The consequent models should contain the information necessary to further applications into flow, transport and mechanical modeling. In any case, the sharpness of these resulting models is related to the amount of sampled areas, and collected data, relatively to the investigated site dimension. Therefore "best" is intended as "best with regards to available data". Such modeling objectives are close to the objective pursued during the SKB site investigations and Site Descriptive Modeling (SDM) efforts performed for both the Forsmark (Olofsson et al. 2007) and Laxemar sites (La Pointe et al. 2008, Wahlgren et al. 2008). But the present approach primarily focuses on the statistical properties of DFN geometry and on their spatial variability, whereas the SDM modeling encompasses firstly deterministic-like information to define Fracture Domains (structural, geological inputs). The method presented here is complementary to SDM purposes.

Fracture networks arise from the combination between fracturing processes, site conditions and determinants like geological factors, stress history, major faults, depth/pressure or rock mass properties. Changes of in situ conditions likely induce changes into the fracture network geometrical properties. But the fracturing process itself induces consequent spatial variability which is not easily distinguished from the former. In this project, we aim at identifying this difference in the observed variability: intrinsic to the fracturing process or extrinsic, that is, constrained by in situ conditions changes. Characterizing the spatial variability of fracturing properties is therefore essential to the whole approach. At the end, it comes to identify, in an objective way, the locations (and relative datasets) for which fracturing is similar (in the sense that it can be reproduced by a unique model and associated parameters) and to define the related models. An additional level of complexity is that the fracturing properties themselves are defined by statistical figures: the fracture density defines the averaged occurrence of fractures into a given spatial domain. Any measure is thus associated with some statistical indetermination of the measured figure, which should be taken into account as soon as datasets and measures are compared.

The report is organized as follows: objectives, limitations and basic definitions are recalled in Chapters 2 and 3. Then, before reconsidering the DFN modeling and classification approach of real cases, we put first emphasis on the notion of variability (variable component of the underlying fracturing process) and on its consequences on the modeling. In this perspective, we initially consider the variability from known stochastic models, in particular Poisson models (Section 4.1 to 4.3) and fractal models (Section 4.4). These models are then used as references when the method is applied to real conditions. The classification method is recalled and illustrated in Chapter 5. Finally the approach, including the variability description of real data, is applied to the SKB Forsmark site and reported in Chapter 6.

2   Approach objectives and limitations

2.1   Key issues

A key element of the methodology is the determination of a rationale for assessing whether local fracture density measurements are statistically similar or not. If they are then they belong to a single group whose associated model characterization arises from the combination of its set of compatible datasets. The related model is called a Statistical Fracture Model (SFM). One main step of the approach is thus to infer the properties of SFMs from which local measurements derive. When datasets and density measurements are not statistically similar, then several different SFMs must be defined.

The other major challenge comes from the consideration of the spatial variability inherent to fracturing properties. The distribution of fractures through space is viewed as the result of a stochastic and variable process. A SFM is defined by a stochastic DFN model which defines some spatial variability. But this one has to be considered together with some statistical indetermination of density measures and also with larger scales spatial variations (Section 2.3) and occurrence of various SFMs throughout a site.

2.2   Practical conditions

The approach is currently developed for **core logging data.** In principle, core-logging datasets are made by a continuous sampling of at least fracture positions and orientations associated to borehole trend, plunge and acute angle. The DFN models are thus density/orientation models and sizes are not included or simply assumed. For boreholes long enough, a complete dataset likely crosses one to several different fracture domains, so that the initial dataset is subdivided into pieces. The initial borehole datasets are divided according to SKB's classification, presented in SDM version 2.0, into Deformation Zones (DZ) and Fracture Domains (FD).

2.3   In situ determinants

There are some deterministic elements which likely affect the DFN organization. For instance, not all the units of a site have undergone the same tectonic events and may therefore display various preferential orientations. Different lithologies may affect the fracture growth rate. The proximity to large shear bands (DZ) likely locally affects the fracture density. Near-surface effects or depth dependent processes may affect the fracture system through depth (Section 6.6).

If these determinants have significant consequences on the fracturing properties the current approach outcomes should reflect these characteristics and therefore reinforce the geological modeling. On the reverse, it is tricky to use some knowledge on in situ determinants upstream to the classification process without entailing some biases in the statistical analysis. One can assume that a set of datasets belongs to a single class and then derive the equivalent model parameters, including possibly a class internal variability larger than class to class properties.

Also, following usual geological based modeling, one would expect some spatial consistency between datasets of a class. While it would be tempting to combine the statistical comparison between density models to a notion of spatial proximity between datasets, it is in practice not simply possible without creating biases. Indeed, imposing an increase of the similarity probability between two close datasets would sometimes be consistent and sometimes inconsistent without any capacity to discriminate; when the spatial boundaries of SFMs are unknown, the probability that two datasets belong to the same SFM is independent of the spatial distance between the datasets. Thus defining physical limits between SFMs and spatial interpolations occur after the classification, not simultaneously.

Note that, one way to indirectly introduce the spatial "continuity" is to start analyses from one dataset and location and then gradually perform the classification by including datasets only below a certain spatial distance. Then one can consider how the possible SFMs first locally defined from few "concentrated" datasets expand towards the all site.

However, if any geological determinant entails a correlation between spatial position and density, the own definition of Fracture Domains and/or SFMs should be reviewed to enable the density variation with some spatial characteristic. Moreover, the current classification algorithm has not the capacity to integrate such a refined rule; but it can for sure currently highlight such a trend.

3   Recalls and basic definitions

A *Statistical Fracture Model (SFM)* is defined by a single DFN model. It reflects the fracturing properties arising from the stochastic nature of fracturing processes together with some in situ conditions. Its characteristics are assessed from a group of compatible datasets (and their respective locations and scales). There is yet no constraint on the spatial distance between the datasets locations.

A *DFN model* is a stochastic model. It defines the statistical distributions of the geometrical properties of a Discrete Fracture Network (DFN). In the present case (i.e. restricted to core-logging data analyses) it encompasses the definition of a *fracture bulk density,* an orientation distribution (could also be the fracture aperture, roughness, etc.), and possibly a model of local spatial variations. Although fracture sizes are part of a complete DFN model, fracture size distributions are not investigated in the present work. We consider that the relation between fracture frequency and fracture bulk density is independent of the fracture size distribution itself.

The *fracture density* is a scale averaging property, which gives the amount of fractures per unit size scale (volume, outcrop surface, or along-core length), with defined properties such as orientations in the present case. In practice the fracture density in 3D is the cumulated surface of fracture by unit of volume (noted d_m). In 2D it is the cumulated length of fracture trace by unit of surface and in 1D it is the *fracture frequency* (noted f, being the number of fracture intercepts by unit of length). The *fracture density* can be defined either as a fully *local property,* that is, when the observation scale tends to zero, or as a *bulk* (macroscopic) one, that is, when the measurement scale tends to "infinity". In general, the fracture density must be related to the length scale over which the averaging is made. The *fracture density of a DFN model* is a bulk density.

A *measure of fracture density* is related to a sample support and a scale. The *apparent fracture frequency f* (also commonly called P_{10}) associated to a borehole section (of size h) is simply:

$$f = \frac{N_{tot}}{h}$$

Equation 3-1

where N_{tot} is the number of discrete fracture intercepts in the dataset.

Similarly, the corresponding *measured fracture density* is given by:

$$d_m = \frac{\sum_{i=1}^{N_{tot}} \psi_i(\hat{a})}{h}$$

Equation 3-2

Where $\psi_i(\hat{a})$ is the orientation bias correction associated to each fracture pole.

In the absence of any other information, one takes the measured density as the best estimate for the DFN model or bulk density. But a *fracture density measure* in fact belongs to a statistical distribution with a *mean (the bulk density)* and a *statistical width* that is likely dependent at least on the number of fractures taken into account in the averaging procedure (the latter being related to the measurement scale). When the measurement scale tends to zero, the underlying distribution has an infinite standard deviation, reflecting that it is not possible to know the fracture density when no fracture is sampled. On the opposite, the larger the sampling scale, the better the fracture bulk density estimate, but the lower the spatial resolution of measurements. The sampling scale is thus a statistical issue in the determination of local values of fracture densities.

Finally, from any initial dataset, both the density estimate and the distribution statistical width constitute the signature of the underlying SFM. The determination of the latter requires quantifying the evolution through scale of the variable part of the fracturing process. How this can be measured in practice is detailed in Section 4.2.

In the following chapter we redefine first the density estimate distribution in case of a commonly used stochastic DFN model, that is the so-called "Poissonian" DFN model. This comes to display how the variable part of the fracturing process determines the density estimate distribution. This reference case is used for further defining the intrinsic variability of SFMs.

4 On the stochastic variability of fracture networks and DFN models

In this chapter we review first how known stochastic DFN models, like the ones based on "Poissonian processes", lead to spatial variation of fracture density and then to density estimate distributions defined by a mean and a standard deviation. Then more complex models, involving a fractal distribution of fractures through space, are considered.

4.1 Case of Poissonian processes to define fracture spatial locations

In the field of probability theory and statistics, "the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time/space if these events occur with a known average rate/density and independently of the time/space since/from the last event" (Haight 1967). A DFN model with random positions of fractures at fixed spatial density thus belongs to the class of Poisson distributions. In the following, we use the terms Poisson (DFN) model or Poissonian model to refer to such models.

In this framework, the fracture density in a volume is viewed as the sum of a large number of independent random variables (fracture in the volume or not), each with finite mean and variance. According to the central limit theorem (Laplace 1814, Moivre 1756), the sum is approximately normally distributed. In the DFN case, the fracture density thus tends to a Gaussian distribution that is only characterized by its mean and standard deviation (*sdev*). *Sdev* is related to the number of summed random variables, which is also the number of fractures observed into the given volume.

Thus the fracture density observed over a given volume belongs to a distribution (mean d_m and *sdev* noted σ_{var}) whose standard deviation σ_{var} reflects the variable part of the fracturing process mimicked by the DFN model (here the Poisson model). The spatial variability of fracture density associated to the Poisson process is therefore characterized by σ_{var} (suffix var recalls the link to variability).

In practice, σ_{var} for a given size of observation should be measured from datasets from volumes sampled at this given size. However, one needs to associate a density estimate and σ_{var} for each individual dataset. Therefore, σ_{var} has to be extrapolated from a scaling analysis of the fracture spatial variability within the dataset. Moreover, we consider a DFN model defined in 3D and sampled in 1D (equivalent to core logging), thus encompassing an orientation bias correction term.

In the following we first define the variability $-\sigma_{var}-$ of the process associated to the fracture frequency (orientation term neglected or specific case with fractures perpendicular to the logging borehole) and finally add the orientation term and consider the density. When frequency, resp. density, is considered σ_{var} is noted σ_f, resp. σ_d.

The characterization of σ_f and σ_d is performed from a probabilistic approach.

Let X (Equation 4-1) be the event for a fracture intercepting a borehole of length L to belong also to a subsection s of L (Figure 4-1). Either the fracture does not belong to s, so that $\psi_0=0$, or it does and $\psi_1=1$. Moreover, if the fracture contribution to the 3D density distribution is accounted for, then $\psi_1=1/(\cos(\alpha_c))$ to handle the orientation bias correction, with α_c defined in Figure 4-1b. The event X is thus defined by:

$$X = \begin{cases} \psi_0 = 0 \\ \psi_1 = 1 \text{ or } \dfrac{1}{\cos(\alpha_c)} \end{cases}$$
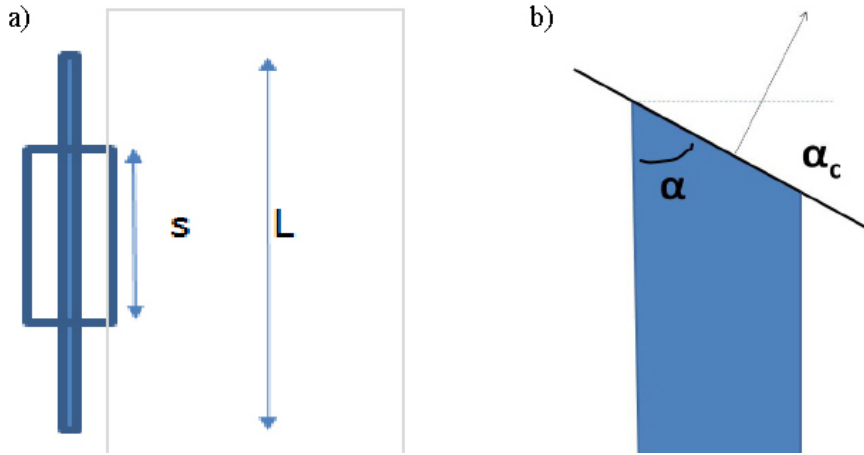
Equation 4-1

Figure 4-1. *a) Schematic view of a borehole section of length L and a smaller scale size s. b) Schematic view of the acute angle α, and angle $\alpha_c = \pi/2 - \alpha$, between a borehole and a fracture.*

Let f (Equation 3-1) be the fracture frequency of the borehole, with $N = f \cdot L$ the total number of fractures sampled by the borehole. The probability p to have one of the N fractures of L also in the sub-section s simply is:

$$p = \frac{s}{L}$$

Equation 4-2

X has 2 possibilities, either it belongs to s or not. Thus the mean $<X>$ and variance $\sigma^2(X)$ associated to X are:

$$\langle X \rangle = \psi_1 \cdot p + \psi_0 \cdot (1 - p) = \psi_1 p$$

$$\langle X^2 \rangle = \psi_1^2 \cdot p + \psi_0^2 \cdot (1 - p) = \psi_1^2 \cdot p$$

Equation 4-3

$$\sigma^2(X) = \langle X^2 \rangle - \langle X \rangle^2 = \psi_1^2 p \cdot (1 - p)$$

X is the event associated to one fracture. When the contribution, to the sub-section s, of all the N fractures of L is considered, we get:

$$\sigma_N^2(s) = \sum_N \sigma^2(X)$$

Equation 4-4

Combining Equation 4-2, Equation 4-3 and Equation 4-4 leads to:

$$\sigma_N^2(s, L) = \frac{s}{L}\left(1 - \frac{s}{L}\right) \cdot \sum_N \psi_1^2(i)$$

Equation 4-5

Since $\psi_1 = 1$ if frequency is considered we have:

$$\sum_N \psi_1^2(i) = N = f \cdot L$$

Equation 4-6

Finally, the standard deviation $\sigma_f(s)$, associated to the fracture frequency at observation size s, is defined by:

$$\sigma_f(s, L) = \frac{\sigma_N}{s} = \sqrt{\frac{f}{s}} \cdot \sqrt{\left(1 - \frac{s}{L}\right)}$$

Equation 4-7

$$\sigma_f(s) \underset{s \ll L}{\longrightarrow} \sqrt{\frac{f}{s}}$$

$\sigma_f(s,L)$ is defined by two terms, one for the finite size effects which becomes dominant when s is close to L and one term in power-law with $s^{-0,5}$. When s is equal to L, $\sigma_f(s=L,L)$ is equal to zero: there is no apparent variability in the frequency assessment at the borehole section scale. But the underlying process variability $\sigma_f(s=L)$ observed a size L is not null and is given by the first term of Equation 4-7. Therefore the density variability of the process has to be corrected from finite size effects and extrapolated as illustrated in Figure 4-2. In the figure, the process variability, in $(f/s)^{-0,5}$ is plotted in black whereas the apparent measures based on finite size samples with L=50 and L=400 are plotted in blue and green respectively.

The preceding development shows that, for a Poissonian DFN model, the frequency estimate standard deviation decreases with the observation scale like:

$$\sigma_f(s) = \sqrt{\frac{f}{s}} = \frac{\sqrt{N}}{s} \qquad \text{Equation 4-8}$$

Thus the standard deviation increases with the number of fractures, but slower than the density estimate which increases like N, resulting in a decrease of the coefficient of variation in $N^{-0,5}$.

The developments from Equation 4-4 to Equation 4-8 are related to a given fracture frequency f and total number of intercepted fractures N. For any real or simulated dataset of a single stochastic process, the frequency will obviously vary from dataset to dataset, simply because of Equation 4-8. In most cases it will fall within the range $[N_{mean} - 2\sqrt{N_{mean}}; N_{mean}+2\sqrt{N_{mean}}]$, where N_{mean} is the mean number of fractures observed at this size. It follows that even if the variability scaling is well described, any dataset variability estimate will be affected by the apparent frequency. This is further illustrated in Figure 4-3: a DFN model (mean density and standard deviation at a given size) is considered at a scale such that N is on average equal to 300. The corresponding model distribution is plotted in a dashed line. Then the two most extreme realizations, after 200 runs are analyzed, result in the "low range" and "large range" plotted in solid lines.

The reasoning to derive the standard deviation associated to fracture density $\sigma_d(s)$ (in m²/m³), instead of the fracture frequency $\sigma_f(s)$ (in m⁻¹) is similar and simply changes by considering ψ_1 as defined in Equation 4-1 instead of equal to 1. The density estimate standard deviation becomes:

$$\sigma_d(s) = \sqrt{\frac{1}{s}} \cdot \sqrt{\left(1 - \frac{s}{L}\right)} \cdot \sqrt{\frac{1}{L} \sum_N \psi_1^2(i)} \qquad \text{Equation 4-9}$$

The evolution with scale of the standard deviation estimate are similar for both the frequency (Equation 4-7) and the density (Equation 4-9), with the term in $\sqrt{(1/s)} \cdot \sqrt{(1-s/L)}$. For the density, the contribution of highly biased intercepts (and thus having a significant weight to compensate the orientation) is dominating with the term in $\psi^2(i)$ in Equation 4-9. The standard deviation in the estimated distribution is larger when the observation is unlikely but potentially highly corrected.
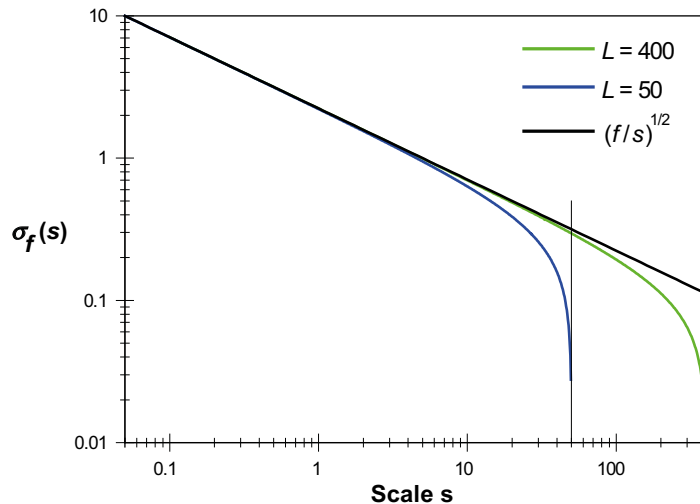

Figure 4-2. *Variability of the density estimate expressed by the evolution of σ_f, see text for details.*
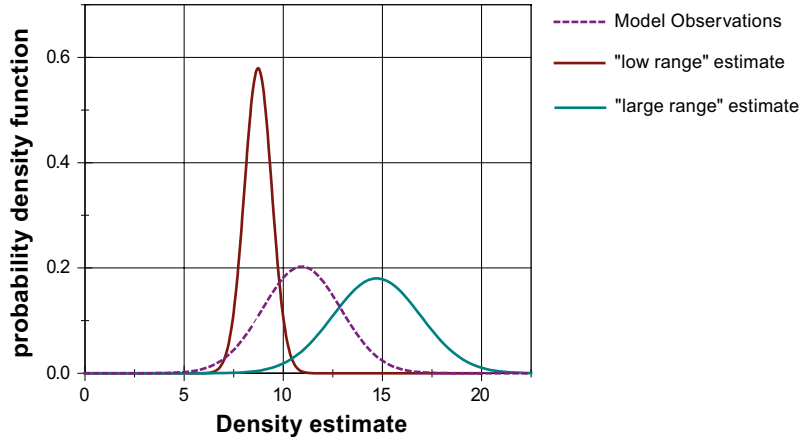
Figure 4-3. *Density distribution model (dashed lines) expected in conditions such that N=300 on average. Corresponding extreme estimates (straight lines) selected out of 200 realizations.*

In summary, in the case of Poissonian DFN models we have:

- A coefficient of variation of the density estimate distribution, expressed by a ratio $\sigma_f(s)/\langle f \rangle$ in frequency, varying like $f^{-0.5}$ or $N^{-0.5}$. Therefore, the largest the sampling volume (or size), the most precise the density estimate and the lowest the standard deviation on the density estimate.

- From one observation (dataset) one can only partly infer the underlying DFN model. The best estimate that can be done is composed by the observed density and its standard deviation σ_{var} (as illustrated in Figure 4-2).

- The Poissonian reference can be used to define guidelines:

  - Equation 4-7 (for frequencies) or Equation 4-9 (for densities) can be used to assess the sample size necessary to reach a certain value of the coefficient of variation of the density distribution estimates:

$$L = \frac{1}{f \cdot \left(\dfrac{\sigma_f(L)}{f} \right)^2}$$
Equation 4-10

  - If the coefficient of variation is 10%, the sampling section size L must be larger than $100/f$. The value increases when accounting for the orientation bias. Designing the sampling conditions to reach certain accuracy in the density estimates fully depends on the underlying DFN model too. Indeed, the total number of sampled fractures, which constrain the variability, both depends on the sample size and density behind.

4.2   Measurement of variability

The variability term of a dataset, σ_{var}, is extrapolated (Figure 4-2), at the sampled section size, from the density standard deviation evaluated at smaller scales. In practice, the sampled section is divided into N_w sub-sections of size s, where each sub-sample is viewed as a smaller realization of the same underlying process. The density standard deviation is thus assessed for varying values of the sub-scale s, as defined in Equation 4-11:

$$\overline{\left(d_m(s)\right)^2} = \frac{1}{N_w} \sum_{i=1}^{N_w} \left(d_{m,i}(s)\right)^2$$

$$\overline{d_m(s)} = \frac{1}{N_w} \sum_{i=1}^{N_w} \left[d_{m,i}(s)\right]$$
Equation 4-11

$$\sigma_{var}(s) = \sqrt{\overline{\left(d_m(s)\right)^2} - \left(\overline{d_m(s)}\right)^2}$$

Finally, $\sigma_{var}(L)$ has to be extrapolated from the evolution of $\sigma_{var}(s)$ at small sizes ($s<<L$), simply because the sample finite size apparently reduces the estimate variability when the sub-sample size s increases. This is illustrated in Figure 4-2.

In practice the extrapolation is done automatically by fitting a power-law (like Equation 4-12) over a restricted range of scales ($s<<L$) and by getting the fitted power-law value at the sample size (L). The power-law exponent thus measured is called a variability scaling exponent, noted β (or β_i, subscript i to denote a particular realization or dataset).

$$\sigma_{var}(s) \sim s^{-\beta} \qquad\qquad \text{Equation 4-12}$$

4.3 Illustration and checking for the Poisson case

Relations on σ_{var} derived in the previous sections are further checked and illustrated for the case of Poissonian DFN models. For this purpose, three Poissonian DFN models (listed in Table 4-1) are defined. They are also used as references throughout the whole study. Fractures are disc-shaped and of constant size (1 m). The size distribution is arbitrary. We recall that this assumption has no consequences when evaluating a Poissonian model orientation distribution and density from 1D or line sampling.

The model VarA has constant orientations; it can be used to generate datasets without sampling bias and therefore frequency equal to density. The model VarB has uniform orientations, i.e. no preferential orientations. Finally, the model VarC has two preferential orientations modeled by two Fisher distributions.

Realizations of the three models are generated into parallelepiped volumes of side 61 m and height 1 m. Sampling 1D lines are simultaneously defined in the horizontal plane with trends equal to: 0°, 29°, 90° and 120°, and centered in (0,0,0). Since fracture sizes are equal to 1 m, the vertical extension of the generation volume around the sampling lines can be limited to 1 m. Sections of length 15 m, 30 m and 60 m are defined. 200 realizations of each DFN model are generated and sampled.

We first consider the scale evolution, in $s^{-0,5}$, of $\sigma_f(s)$ and $\sigma_d(s)$. It is illustrated for model VarA in Figure 4-4. The evolution of $\sigma_d(s)$ is averaged over 200 datasets of section length equal to 60 m. In this case, the average number of fracture intercepts on one dataset is equal to 600. The scale evolution in $s^{-0,5}$ is well observed (black line in Figure 4-4). The finite size effect (term in s/L in Equation 4-9) is visible from s equal to 3–4 m (for a section length $L=60$ m).

Moreover, as shown in Figure 4-4, there is a discrepancy between the expected relation defined in Equation 4-9 and the simulations. The observed finite size effect term is in $(1-s/L)$ instead of $(1-s/L)^{0,5}$. This shift is observed for all the cases referenced in Table 4-1. The discrepancy is only on the finite size effect term. This is likely due to conditions of the sampling simulations (here fractures are generated in 3D with constant sizes and sampled along 1D lines). When straightaway randomly generated along a 1D line, the fracture intercept positions variability $\sigma_d(s)$ follows Equation 4-9 with the term in $(1-s/L)^{0,5}$. Since the scale independent term in $s^{-0.5}$ is not affected, at that stage, this is not further investigated.

Table 4-1. Poissonian DFN models.

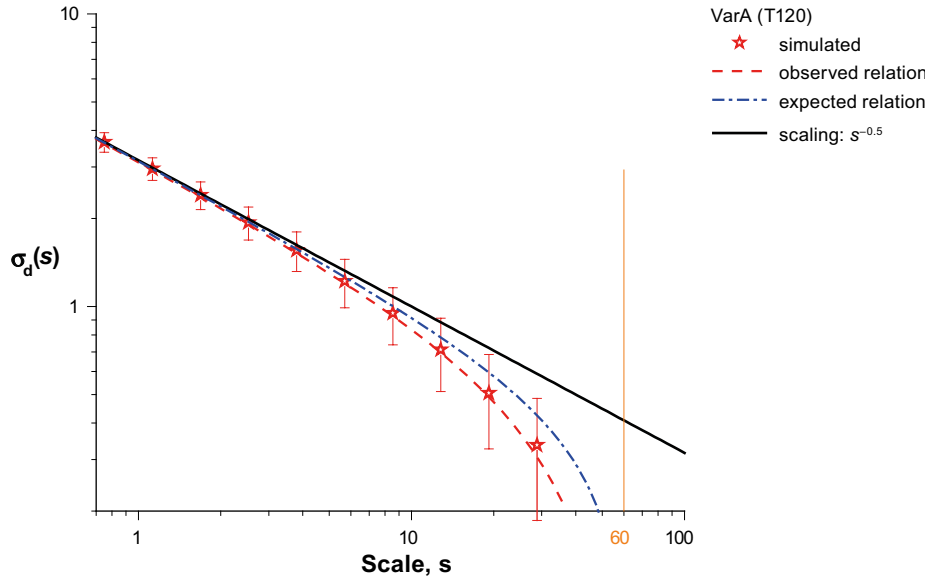| Name	Density d_m m²/m³	Orientations	Number of generated realizations
VarA	10	Constant (strike 30, dip 90)	200
VarB	10	Uniform (no preferential direction)	200
VarC	10	2 Fisher distributions: dip ; dip direction ; kappa 90 ; 23 ; 10 0 ; 10 ; 10	200

Figure 4-4. *Evolution of <σ_d(s)> for the DFN model VarA, averaged over 200 datasets of section length equal to 60 m and resulting from the 1D sampling in direction trend 120.*

The observed evolution of $\sigma_d(s)$ finally is defined by:

$$\sigma_d(s) = \sqrt{\frac{1}{s} \cdot \left(1 - \frac{s}{L}\right)} \cdot \sqrt{\frac{1}{L} \sum_N \psi_1^2(i)}$$

Equation 4-13

In Figure 4-5 we illustrate how orientations (third term in Equation 4-13) affect the variability σ_d. All fractures of VarA are vertical and striking 30°, so that the orientation bias is null for a horizontal borehole trending 120° (noted T120) and equal to 1/cos (89°) for a borehole trending 29° (noted T29). With uniform orientations in VarB, the average sampling bias is independent of borehole orientation. For a sampling section length equal to 60 m, the VarA-T120 datasets contain on average 600 intercepts, whereas the VarA-T29 have only about 11 intercepts and the VarB ones have 300 intercepts.

The extrapolated value of $\sigma_d(s)$ at the sampling size 60 m for VarA and borehole direction T120 is equal to 0.4, so that the ratio $\sigma_d(s)/d$, with d the density, is equal to 4%. If the borehole orientation changes so that orientation bias increases and the number of intercepted fractures decreases, the ratio can increase up to 30% if the acute angle is 1° (which is unlikely to happen in reality). For the uniform distribution of fracture orientations (VarB), the ratio is about 6% (when the average number of fractures in a dataset is 300). These values are mainly linked to the real number of fractures into a dataset (so linked to the parent density and borehole length). Also, variations observed from one realization to the other are larger when sample bias is more important (see the sdev bars on the standard deviation estimates in Figure 4-5 and histogram of the exponents in Figure 4-6).

Preceding figures show that on average, the variability term $\sigma_d(s)$ is well defined by Equation 4-13. There are however important variations from realization to realization. This is illustrated first through the histogram of estimated variability scaling exponents β_i computed from individual realizations i. Basically β_i is automatically measured from all the simulations of 60 m length datasets of VarA and VarB (respectively in direction T120 and EW). Corresponding histograms of the exponent absolute values are plotted in Figure 4-6. The variability exponents are distributed in the range [0.35; 0.75], with most values close to 0.5. Both VarA and VarB histograms are similar, with a mean value close to but slightly larger than 0.5. This shift is likely linked to the automated fitting procedure combination with finite size effects.
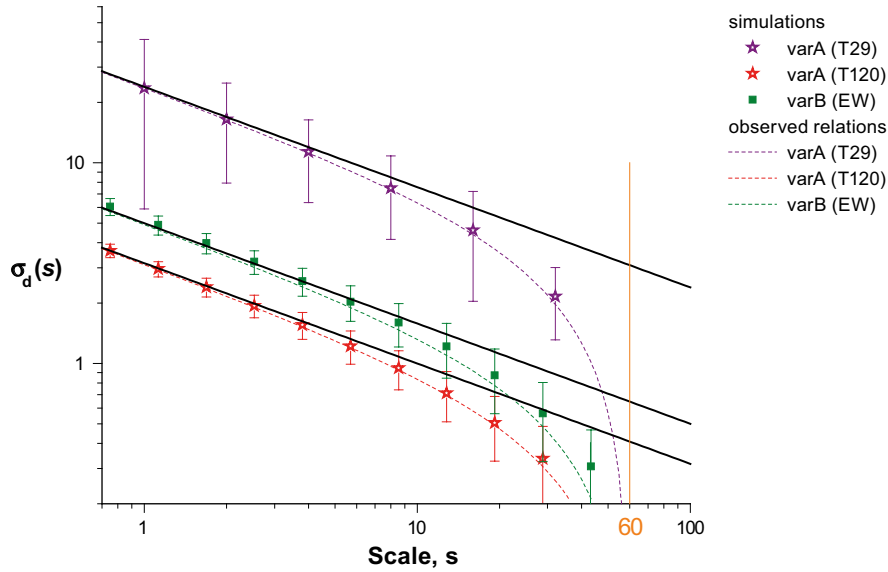
Figure 4-5. *Evolution of $<\sigma_d(s)>$ for the DFN models VarA (2 sampling directions, T29 and T120; sampling length equal to 60 m) and VarB (EW direction). Simulated data are plotted as symbols, observed relation (Equation 4-13) as dashed lines (see inlet for details) and scaling in $s^{-0.5}$ in solid black lines.*
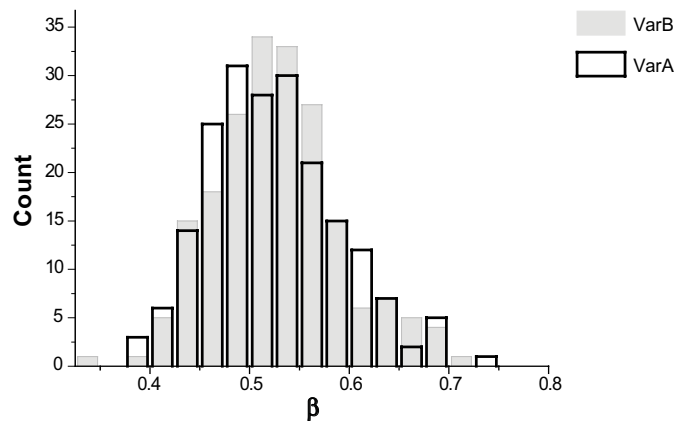
Figure 4-6. *Histogram of the measured exponents β_i for 200 realizations of models VarA and VarB.*

Moreover, there is no apparent correlation between β_i and the apparent density assessed for each realization. We finally highlight, in Figure 4-7, the correlation, by realization i, between $\sigma_i(h)$ and β_i, with h the extrapolation length. For each dataset, the value of $\sigma_i(h)$ is normalized by the equivalent variability term, $\sigma_P(h)$, directly computed from Equation 4-13.

Once normalized by $\sigma_P(h)$, which reflects the density term, the variability term extrapolated at the section length is similar for the VarA and VarB cases. It depends on β_i through an exponential correlation. Therefore the variability term $\sigma_i(h)$ is larger when β_i is low.

Even for datasets having up to 600 fracture intercepts (case VarA at section length 60 m), the Poissonian model displays sometimes apparent correlations (β_i different from 0.5) departing from the averaged process.

In the following sections we consider non-Poissonian models.

Figure 4-7. *Observed evolution of $\sigma_i(h)/\sigma_P(h)$ as function of β_i for 200 datasets (length 60 m) for both VarA and VarB. $\sigma_P(h)$ is computed with Equation 4-13 for each dataset (the total number of fracture intercepts varies from dataset to dataset). Results for VarA (resp. VarB) are plotted in green (resp. blue) symbols and exponential fit in black line.*

4.4 Case of a stochastic fractal model to define fractures spatial locations
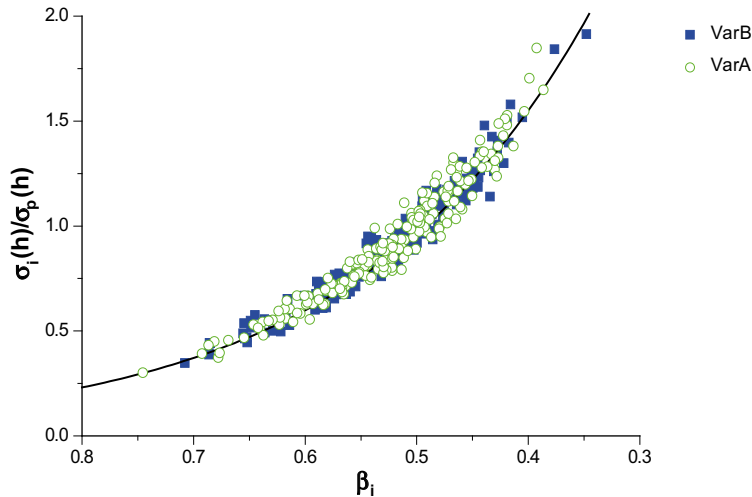
In addition to the above studied uniform Poisson DFN models, DFN models with a fractal structure of fracture positions are worth being considered in parallel to natural systems. They indeed display specific long-range correlations which likely induce significant scaling effects. Although some fault networks have been found to be characterized by fractal dimensions significantly lower than topological dimensions (Bonnet et al. 2001), the fracture networks of the SKB sites (Darcel et al. 2004, 2006) are characterized by relatively weak fractal dimensions (i.e. close to Euclidean dimension).

The results presented here are preliminary.

4.4.1 The 1D fractal case

We first describe the evolution of $\sigma_f(s)$ from a 1D fractal DFN model. Fracture positions of the model have a fixed predefined fractal dimension. The fractal dimension ($D \leq 1$) is generated from a multiplicative cascade process (Darcel et al. 2003b) similar to a Cantor process in 1D. Thus the fractal dimension of a simulated dataset is directly the dimension D.

The simulated datasets are in the range [30 m; 480 m] and contain on average 5 fractures per meter. Exploratory analyses were performed for a range of fractal dimensions. We present preliminary results with $D=0.98$ and show that even for a very weak fractal dimension the variability is significantly affected. The resulting fracture frequency variations are illustrated in Figure 4-8 below, for $D=0.98$ and 480 m. Although the fractal dimension is "weak", we observe long range variations which are unlikely to occur with Poissonian DFN models.

The evolution of $\sigma_f(s)_L$ is averaged over a large number of realizations. It differs on average from a Poissonian variation (Figure 4-9a) with a slower decrease with size. The derived β exponents are comprised between 0.3 and 0.4. Also we observe limited variations with the domain size L (Figure 4-9a). The larger L, the larger is the variability at s fixed.

More investigations are necessary to relate exactly the variability exponent to the fractal dimension. The presented simulations nevertheless are complementary to the previous Poissonian case to determine what model of variability could be consistent with real data.
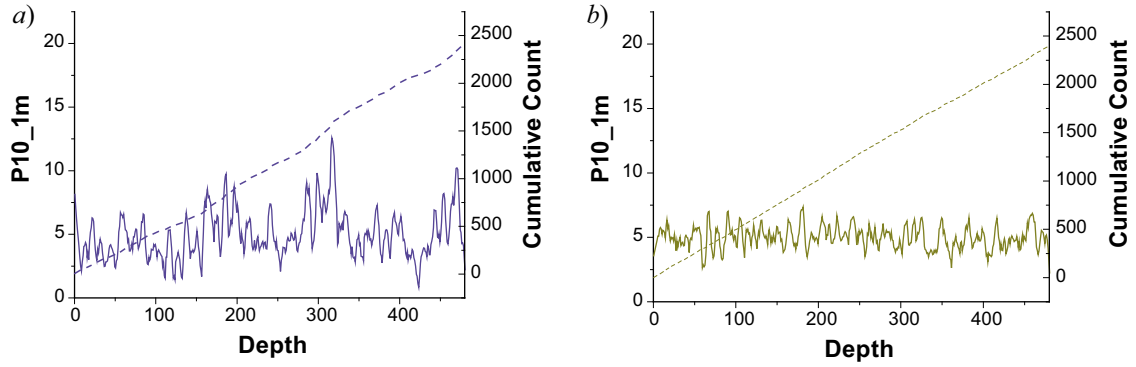
Figure 4-8. *Fracture frequency by meter, noted P10_m, (left vertical axis continuous line) and cumulative count (right vertical axis and dashed line in the plot) as function of position or depth along a linear sampling section for a) Cantor dataset with D=0.98 and b) Poissonian dataset.*
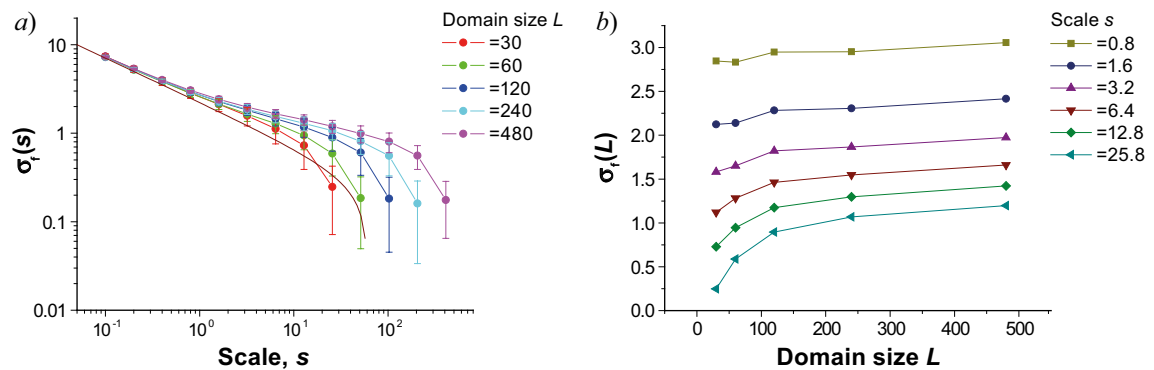

Figure 4-9. *Case D=0.98. a) Evolution of σ_f with scale s for several samples of lengths L. The brown solid line is the equivalent Poisson model for L=60 m. b) Evolution of σ_f (L), i.e. with sample size L, at constant observation scale s.*

4.4.2 The 3D fractal case

In complement to the section above, we consider a DFN model whose properties (including the fractal dimension of fracture positions) are defined and generated in 3D directly and then sampled in 1D. We take a distribution that is likely close to natural fracture networks with both a fractal distribution of fracture centers and a power-law size distribution of fractures. In such case the fractal dimension observed from the fracture intercept positions can be partly hidden by the fracture size distribution (Darcel et al. 2003a).

Two fractal dimensions are considered, equal to 2.5 or 2.95, and two power-law size distributions with exponents *a* equal to 3.1 or 4. The 4 models thus defined are associated to a density equal to 15 (m^2/m^3). In each case, 4 realizations are generated into a $100 \times 100 \times 100$ m volume and sampled according to boreholes of size 10 to 100 m.

The resulting variability exponents are plotted in Figure 4-10. The largest variability is observed for the smallest fractal dimension (*D*=2.5), with a variability exponent *β* close to 0.1. When the fractal correlation decreases (*D*=2.95) the variability exponent increases up to almost 0.5. When the size exponent decreases from 4 to 3.1, *β* increases from 0.1 to 0.2 (for *D*=2.5): a large proportion of large fractures tend to hide the fractal structure when observed from 1D line sampling (Darcel et al. 2003a). *β* values higher than 0.5 occur only for the smallest datasets (mainly *L*=10 and about 150 data by dataset), likely due to the too small size of the dataset. Beyond 50 m and about 750 fracture intercepts, the variability exponents are well defined (Figure 4-10).

Similarly to the 1D fractal case (previous section), even a weak fractal correlation in 3D leads to a change of the variability exponent *β* (by comparison to the non-fractal Poisson models from Figure 4-6). At the extreme, a strong fractal correlation entails a large increase of the standard deviation associated to the density estimate.
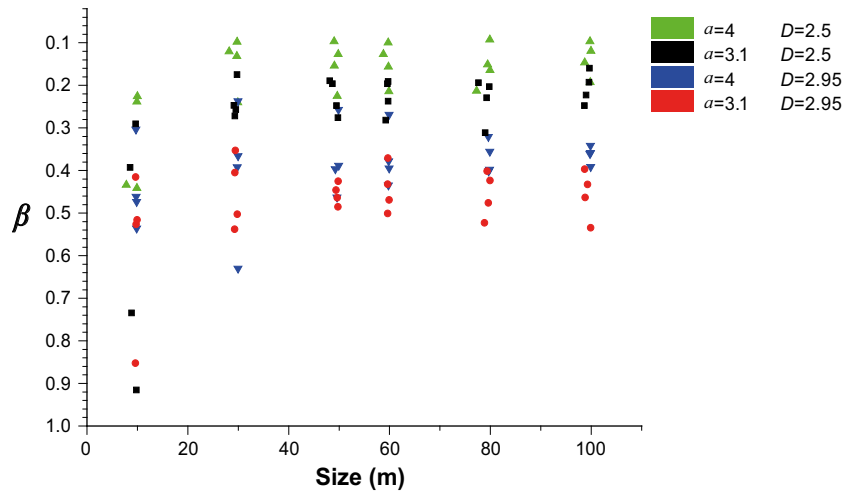
Figure 4-10. *Variability exponents for fractal DFN models. See text for details.*

5   Classification of fracture datasets into SFMs

With the classification process we aim to identify the best division of density distribution estimates into statistically distinct classes. Each class then refers to a Statistical Fracture Model (SFM). In practice the method identifies which datasets, among an initial set of datasets, display compatible statistical fracturing properties and clusters them accordingly, in order to finally infer the underlying density distribution of their relative SFM. Each dataset is firstly associated with one density distribution estimate and one class; then the classification algorithm is pursued until a certain acceptable variability threshold is reached into each defined class and finally the SFMs are defined. The methodology was initially presented in Darcel et al. (2009) and further illustrated in Darcel et al. (2012). In this chapter we recall the rules of the method and illustrate it with an application to synthetic data from Poissonian DFN models.

5.1   Method

The difference, noted $\Delta(j,k)$, between two density distribution estimates (datasets j and k) is defined by the absolute difference between the two corresponding densities d_m, normalized by the sum of their standard deviations σ_{var}. If density distribution estimates are discretized according to orientation bins ($n=1$ to N_Z orientation bins, see Figure 5-1), then the average difference, $\Delta_a(j,k)$, is equal to:

$$\Delta_a \left(j,k \right) = \frac{1}{N_Z} \sum_{n=1}^{N_Z} \frac{\left| \left(d_m \right)_n^j - \left(d_m \right)_n^k \right|}{\left(\sigma_{var} \right)_n^j + \left(\sigma_{var} \right)_n^k} \qquad \text{Equation 5-1}$$

$\Delta_a(j,k)$ thus defines the *averaged statistical distance* between any pair of datasets.

With this definition, a small difference at an orientation bin may compensate a larger difference at another orientation bin. A second more restrictive criteria, the maximal statistical distance noted $\Delta_r(j,k)$, is defined to ensure a fulfilled criterion for each orientation bin:

$$\Delta_r \left(j,k \right) = \max_{n=1 \cdots N_Z} \frac{\left| \left(d_m \right)_n^j - \left(d_m \right)_n^k \right|}{\left(\sigma_{var} \right)_n^j + \left(\sigma_{var} \right)_n^k} \qquad \text{Equation 5-2}$$

The use of one criterion or the other is further discussed in Darcel et al. (2012).

The datasets within a class – and their associated distribution estimates – belong to the same SFM, with fracturing properties satisfactorily described by the properties of the class itself. Two different classes reflect two different SFMs. The number of classes defines the number of different models required to fully define the fracturing properties of a site.

The method is derived from the agglomerative hierarchical clustering algorithm of Chelcea et al. (2006). It is based on an automated sorting process which generates disjointed classes that group objects with similar properties. Thus it defines classes with specific characteristics by partitioning the space of all possible models into different subspaces.

The classification process includes any initial amount of datasets. It groups the datasets into several classes and computes the associated equivalent density distributions. A class density estimate (d_{cs}, σ_{cs}), where N_{cs} datasets are grouped, is given by:

$$\left(d_{CS} \right)_n = \sum_{i=1}^{N_{CS}} \frac{\left(d_m \right)_n^i}{\left(\sigma_{var} \right)_n^i} \Bigg/ \sum_{i=1}^{N_{CS}} \frac{1}{\left(\sigma_{var} \right)_n^i} \qquad \text{Equation 5-3}$$

$$\left(\frac{1}{\sigma_{CS}} \right)_n = \frac{1}{N_{CS}} \sum_{i=1}^{N_{CS}} \frac{1}{\left(\sigma_{var} \right)_n^i}$$

To get the final classification, the sorting algorithm proceeds as follow:

- First, a class is defined for each dataset;

- Then, an iterative process starts:
    - The two closest classes are merged together;
    - The existing classes are updated, that is, if a dataset is closest to another class, the class definitions are accordingly changed and class estimates redefined (Equation 5-3);

- The iterative process is stopped when the largest dispersion (from the difference between the class and its datasets, Equation 5-1) within an existing class exceeds a fixed threshold (called Δ_{max}).

$$\Delta_{max} = \max\left\{\Delta_a(j, cs_j)\right\} \qquad\qquad \text{Equation 5-4}$$

In the perspective of the classification algorithm, a unique division into orientation bins must be used at once and simultaneously applied to any amount of datasets. Defining orientation bins is a compromise between the objective (to precisely identify preferential orientations of the fractures) and relative accuracy decrease (many zones imply less data in each one). Thus, when increasing N_z, the relative increase of σ_{var} with regard to d_m automatically decreases the terms of $\Delta(j,k)$ thus leading to limit the capacity to identify classes. Basically, if we subdivide a fracture set into different orientation classes, we also decrease the number of fractures in each subset and thus increase the variability of the underlying statistical distribution. In case of a uniform orientation distribution, the standard deviation increase is exactly defined by Equation 5-5:

$$d_m\big|_j = \frac{d_m}{n_{zones}}$$

$$\sigma_{var}(L)\big|_j = \sqrt{\frac{f}{L \cdot n_{zones}} \cdot \pi} = \sigma_{var}(L)\sqrt{\frac{1}{n_{zones}}} \qquad\qquad \text{Equation 5-5}$$

$$\frac{\sigma_{var}(L)\big|_j}{d_m\big|_j} = \frac{\sigma_{var}(L)}{d_m} \cdot \sqrt{n_{zones}}$$

The definition of orientation bins, as illustrated in Figure 5-1, is an issue. Ideally, it would reflect evidence of preferential orientations (geological fracture sets) but in practice the number of orientation bins has to be limited (see above) and there is no unambiguous such division, especially when the amount of datasets is important. Several simple possibilities are reviewed in Figure 5-1; first the "O-1" grid, with $N_z=1$, simply refers to a unique value of density, the entire density. With the "O-3" grid the stereonet is divided into three dip bins and thus separates subhorizontal fractures, subvertical fractures and intermediate fractures. This grid is adapted to focus on the particular effect of vertical position or depth and relative dip of the fractures. Two possibilities of "O-5" grids, i.e. with $N_z=5$, are also displayed in Figure 5-1c and -d. Whatever the chosen grid, there is only a partial correspondence between clusters of preferential orientations and orientation bins, even in the simplest case where only one dataset (as for dataset KFM08C-FFM01d in Figure 5-1) is considered.

Even if an orientation grid does not fit with the apparent geological sets of a dataset, it does not make the consequent density estimates wrong, but highlights more or less successfully the occurrences of orientation clusters. This is further illustrated with VarB and VarC models (defined in Table 4-1). 5 realizations of the two models are generated and synthetic borehole data generated from a single borehole orientation. Density distribution estimates are computed for the two "O-5" grids of Figure 5-1c and -d and displayed in Figure 5-2a and b. We first note that the value of σ_{var} is increased when the orientation domain corresponds to highly biased values (like in bin 1 and 4 of Figure 5-2a, see Equation 4-9).

Let recall that model VarB is uniformly oriented whereas model VarC is defined by two preferential orientations. The latter is better highlighted in Figure 5-1c than in Figure 5-1b. Also in Figure 5-2b there is likely one preferential orientation of VarC, and therefore cluster of fractures, which is split through 2 orientation bins, leading to less apparent differences between bins 3 and 4. Both models have the same total density, which is defined by the sum of each orientation bin associated density, therefore independently from the chosen grid.

a) "O-1"

b) "O-3"

KFM08C_FFM01D

KFM08C_FFM01D

c) "O-5"

d) "O-5"$_{KFM08C}$
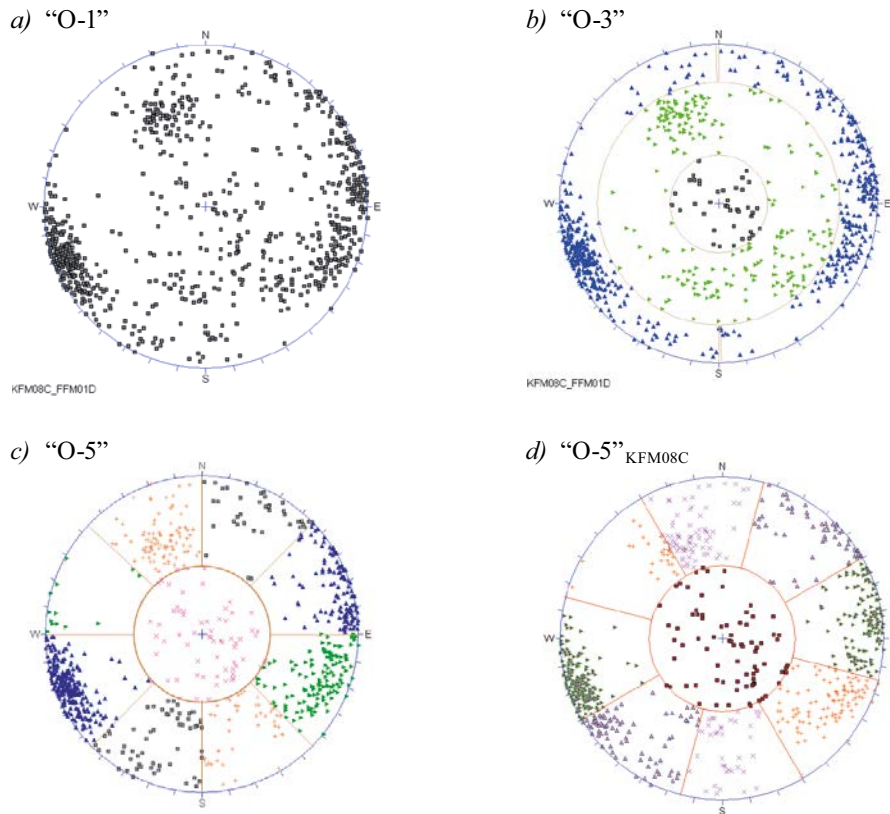
Figure 5-1. *Illustration of dataset orientation binning, from dataset KFM08C-FFM01d. a) grid "O-1": no binning, all orientations together, one value of the whole density, b) grid "O-3": binning into 3 bins referring to vertical direction, i.e. from sub-horizontal (bin 1) to intermediate (bin 2) and sub-vertical (bin 3) fractures but not distinction into strike, c) "O-5" and d) "O-5$_{KFM08C}$" systematic binning into 5 bins, one horizontal and 4 "vertical", according to strike. d) "O-5$_{KFM08C}$" is qualitatively adapted to KFM08C apparent pole cluster, the vertical cells are shifted with regards to case c).*

a)

b)
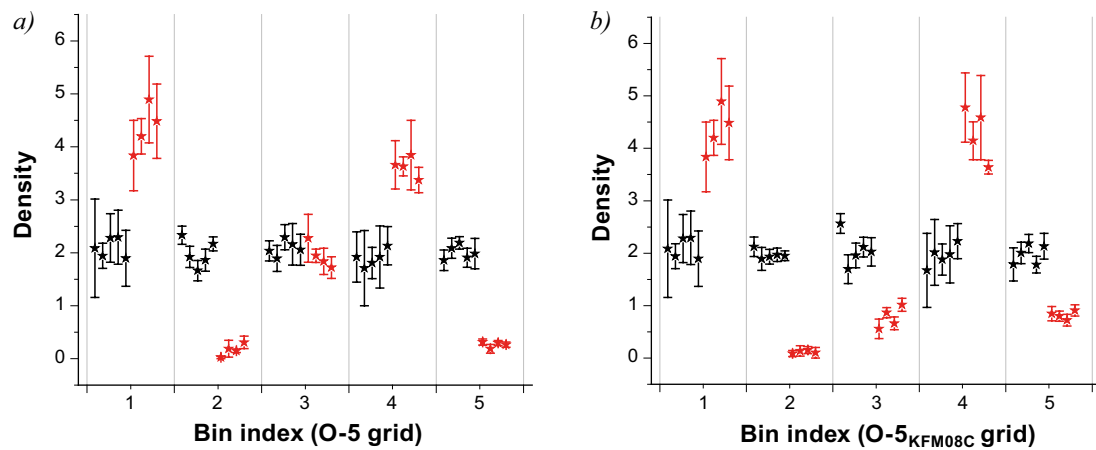
Figure 5-2. *Density estimates from grids similar to Figure 5-1c) and d) for 5 realizations of each model VarB (black) and VarC (red).*

Finally, the classification algorithm ends up into the right class definition (red and black colors in Figure 5-2) for both cases.

In summary, except for simple cases, there is no unique, or "best", division into orientation domains. We assume however that the impact of this division is of secondary importance compared with density variations.

5.2   Application and illustration

The method capacity to differentiate fracturing models, especially the role of Δ_{max}, is illustrated further from the synthetic Poissonian DFN models, varA, varB and varC (Table 4-1).

First the 200 realizations of the DFN model VarB are sampled and analyzed. The average number of fractures intercepts by borehole meter is close to 5. Realizations are sampled along 60 m length boreholes, so that on average the number of discrete data in one dataset is close to 300. The density estimates assessed for the 200 realizations are displayed in Figure 5-3a. The estimate standard deviations are automatically computed.

We apply the classification process until all realizations are grouped together. The corresponding evolution of Δ_{max} is displayed in Figure 5-3b. The required Δ_{max} to get one final single class is equal to 1.66. This value can be taken as a reference to define a threshold for practical applications.

In a second time, we apply the classification process to several realizations (5 by model) from the three reference models (Table 4-1) and test three different orientation grids: the O-1 and O-5 grids of Figure 5-1 and the "O-13" grid defined in Figure 4-1c of Darcel et al. (2009). This last one provides an extreme case with a large number of bins.

Results are plotted in Figure 5-4. The classification is capable of identifying the three different models (with Δ_{max}=1.66) with a perfect accuracy except for the 1-bin grid, since the three models have the same total average density. This value steeply increases if one forces the classification up to less than 3 classes.

5.3   Outcome

At the end of the classification process, a SFM is defined from the combination of several compatible datasets, including a density estimate, a standard deviation and a variability scaling exponent (as schemed in Figure 5-5). The standard deviation on the density estimate, σ_{var}, is related to scale of observation and to the variability scaling of the process.

A SFM provides a stochastic description of the fracturing properties and can be used to generate consistent DFN realizations at a given scale (Figure 5-5).
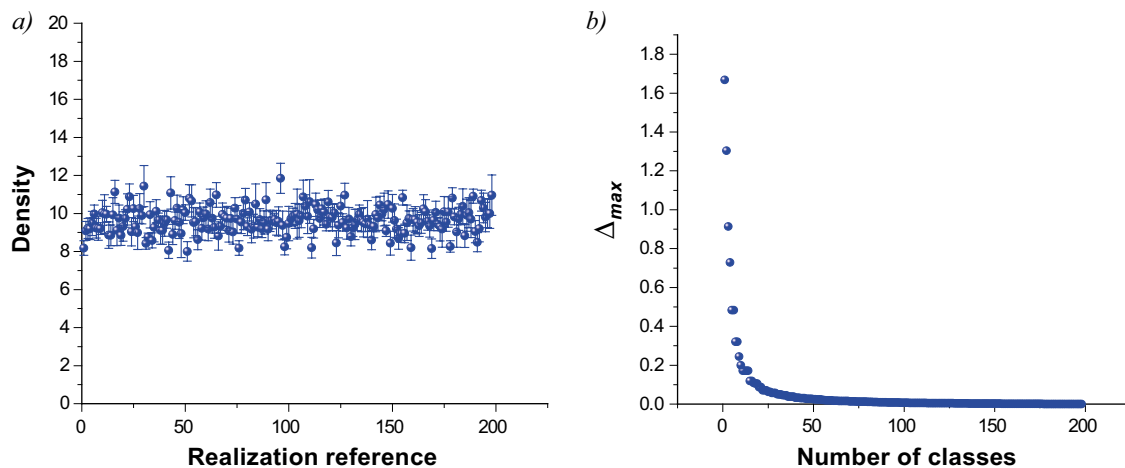
Figure 5-3. *a) Distribution estimates, i.e. mean (filled circles) and standard deviation (bars), for 200 realizations (reference on the x-axis). b) Evolution of the maximum class dispersion as function of the number of classes during the classification process. Here Δ_{max}=1.66 is the final dispersion of the unique class for the 200 realizations of the same underlying process.*
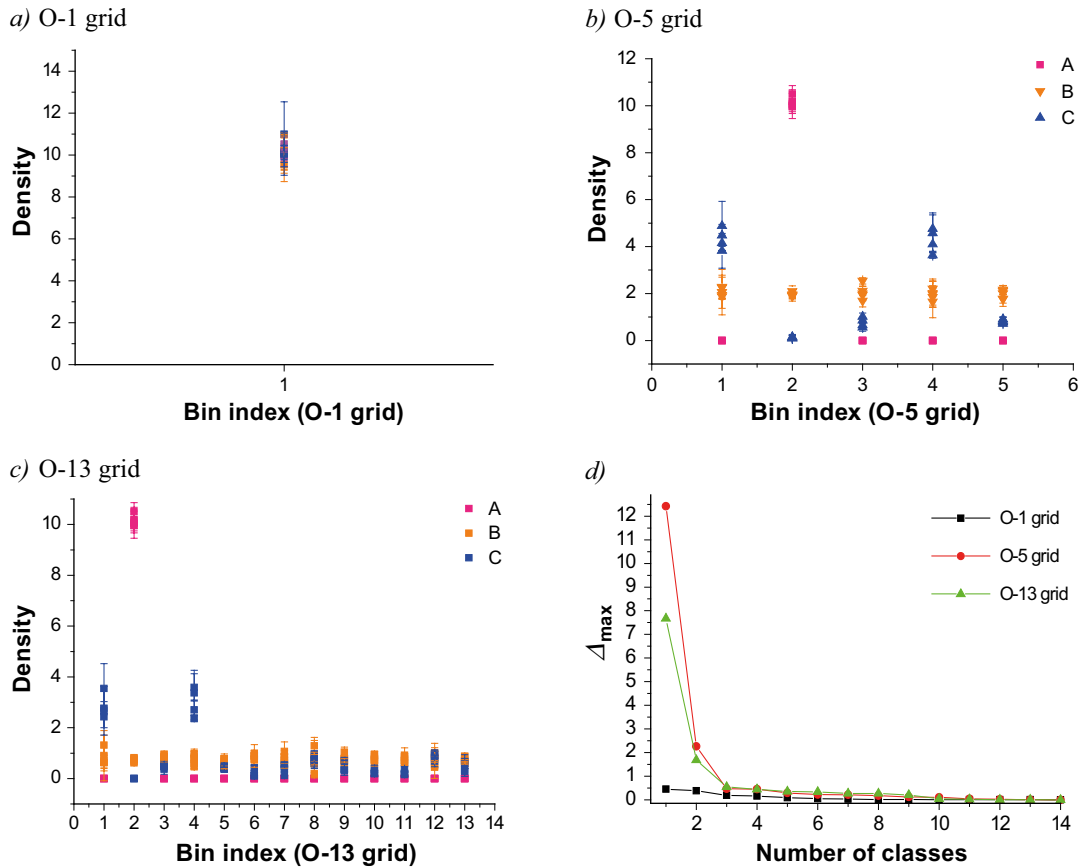
Figure 5-4. *Results of the classification applied on three different Poissonian processes (varA, varB and varC) with three different orientation grids: a) "O-1", b) "O-5" and c) "O-13". Dots are colored according to the final class indexes. d) Evolution of the maximum class dispersion as a function of the number of classes during the classification stage.*
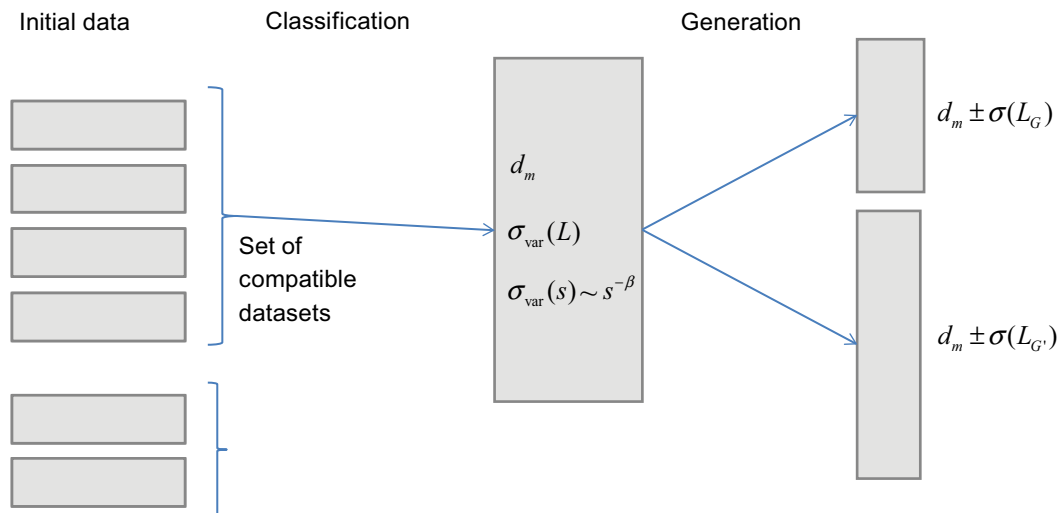
Figure 5-5. *Steps of SFM definition. From left to right, a set of initial datasets is identified as compatible and thus used to define the SFM of the class. This includes d_m the density estimate, $\sigma_{var}(L)$ the standard deviation at scale L, where L refers to the cumulative size of the class datasets, and the variability scaling, through exponent β. Finally, the density estimate standard deviation expected at a given observation size (L_G or $L_{G'}$) is defined by the combination between the variability and scaling and the observation size itself. It naturally decreases when increasing volume of consideration.*

6 The fracture organization in Forsmark

The approach described above is applied to the SKB Forsmark site. The Poisson (and fractal) DFN models are used as reference cases for the analysis of the real data density distribution estimates.

6.1 Database

The selected data are the cored borehole data (Sicada delivery 08_95) combined to interpretations of the Site Descriptive Modeling (SDM) reported in Olofsson et al. (2007): each entire borehole dataset will be divided into smaller sections according to the SDM Fracture Domains (FD, named FFM01 to FFM06 at Forsmark) and Deformation Zones (DZ) of Olofsson et al. (2007). All the defined datasets are listed in Appendix 1: Forsmark datasets.

6.2 The field case, intrinsic variability of fracture density: scaling issues

The longest boreholes give the longest possible continuous sampling of fracture density and variability, likely through several distinct FDs and DZs. In a preliminary step, we consider the density variability at the scale of entire boreholes. This rough definition of "one domain by borehole" is not optimized. It surpasses the FDs and DZs classifications and comes to typically encompass local density variations as large as the ones highlighted in Figure 6-1. The observed combination between small and large scale density variations is likely to end up with a variability scaling larger than what would be expected from a Poisson model (Equation 4-13). Therefore the resulting variability scaling exponents fall in the range [0.2; 0.35], as plotted in Figure 6-2 (brown histogram). This range is below the value (0.5) expected for a Poisson process. Small values of the variability scaling exponent will end up into a larger variability and larger standard deviation associated to the density estimate. If distinct SFMs are combined into one single dataset, one naturally expects more internal variability in the resulting domain and poor understanding of the fracturing determinants.

We then assume that, for each borehole, the SFMs changes are most likely defined by the current DZs and FDs local boundaries. Therefore each time a change in FD and/or DZ is encountered along a borehole (illustrated through a change of color in the base line in Figure 6-1), a new dataset, or borehole section, is started and the preceding ended. Too thin sections (i.e. width less than 2–3 meters) are neglected and merged into their neighboring domain. For the example in Figure 6-1, the final sections are recalled in Table 6-1 below:

Table 6-1. Example of borehole KFM08C: "our" basic datasets, or borehole sections, defined from and existing FD and DZ division.

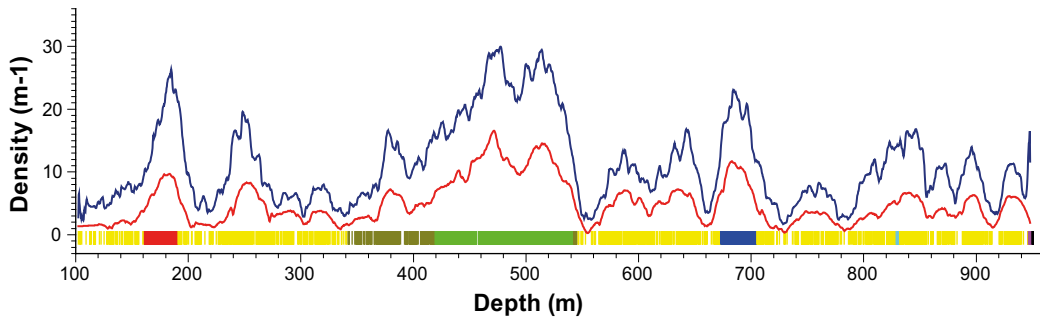| IDCODE	Depth_min	Depth_max	IDCODE_section
KFM08C	102.3	161.1	KFM08C-FFM01a
KFM08C	161.1	190.6	KFM08C-DZ1
KFM08C	190.6	342.2	KFM08C-FFM01b
KFM08C	342.2	419.3	KFM08C-FFM06
KFM08C	419.3	541.8	KFM08C-DZ2
KFM08C	541.8	673.1	KFM08C-FFM01c
KFM08C	673.1	704.9	KFM08C-DZ3
KFM08C	704.9	950.5	KFM08C-FFM01d

Figure 6-1. *Borehole KFM08C, local density (blue line) and frequency (red line) variations. The division into FDs (yellow and brown colors) and DZs (red, green, blue and violet colors) sections is schemed at the figure bottom. The local frequency P_{10} (red line), and d_m (blue line), are smoothed by adjacent averaging over close to 19 m.*
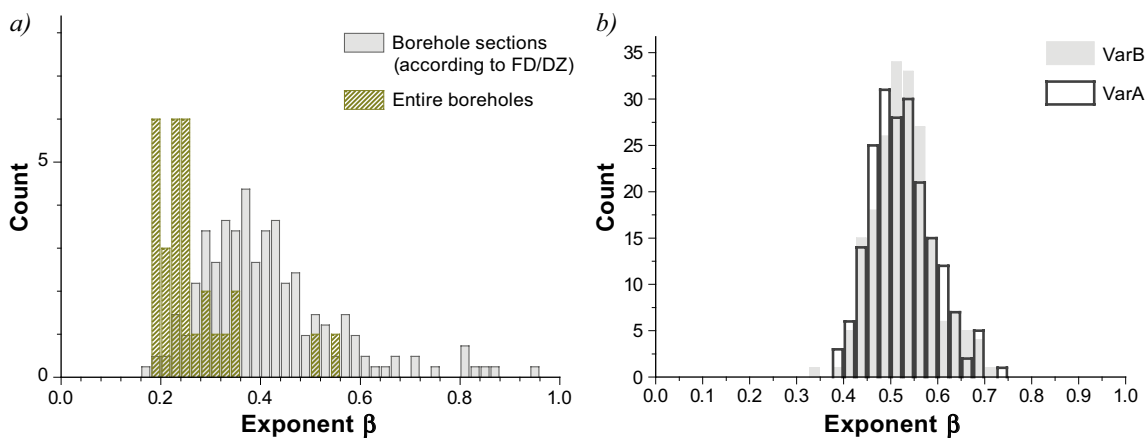
Figure 6-2. *Exponents β (of entire density) histograms for (a) entire boreholes and boreholes segmented according to the FD/DZ division at Forsmark and (b) reference DFN models VarA and VarB.*

Borehole KFM08C is thus divided into 8 datasets. In general one would expect that a DZ section would be thinner than a FD section, since DZs are defined as major shear bands, thus more or less 2D very large planar zones of disturbance. But we observe with KFM08C at least one really large DZ (light green in Figure 6-1 and IDCODE_section KFM08C_DZ2), so that there is no systematic difference in size between the FD and DZ datasets.

All over the Forsmark cored boreholes, the total number of datasets defined is 148. Dataset section sizes are between 5 m and 522 m and the range of number of data in each is between 5 and 4,229. The assessed variability exponents will thus arise from varying sampling conditions in term of segment sizes and number of intercepted fractures. The smaller is the number of data in a dataset (due to low density and/or small size), the larger is the variability; also, the smallest is the segment size, the largest is the bias relative to finite size effects towards small values (see Figure 4-2). Whatever the initial conditions, the variability exponents are automatically extrapolated. The resulting histogram is displayed in Figure 6-2a, whereas the histogram of the Poissonian analyses is plotted in Figure 6-2b.

Like for the Poisson reference models, the real datasets end up into a distribution of scaling exponents *β*, but centered on 0.35 whereas the Poisson models are centered on 0.5. Therefore the variability associated to the borehole sections decreases faster than the variability associated to entire boreholes, but slower than a Poisson process.

The analyses performed on the Poisson models have shown that data sampling leads to apparent clusters and therefore to a distribution of values around the theoretical one rather than to a unique variability exponent. This effect decreases while increasing the amount of data to perform the analysis. By analogy, the observed distribution for the real datasets may arise from a single underlying variability scaling whose process value would be very close to 0.35.

We investigate further the possible origin of the variations towards the datasets size – equivalent to number of elements, belonging to a DZ or a FD. First, Figure 6-3 displays the scaling exponent versus the estimated density, with dots colored according to the number of data by dataset. We observe that the largest datasets are in the range 0.2 to 0.4 for β, whereas values up to 0.8 and over are assigned to weakest datasets.

In Figure 6-4, we show the scaling exponent by distinguishing FDs and DZs and the resulting distribution of scaling exponent for the two types of datasets. We observe no significant differences between FD and DZ in terms of scaling exponent.

The only apparent correlation between single β values and dataset parameters is a concentration of values in the range [0.2; 0.4] when the number of data by dataset is large enough.
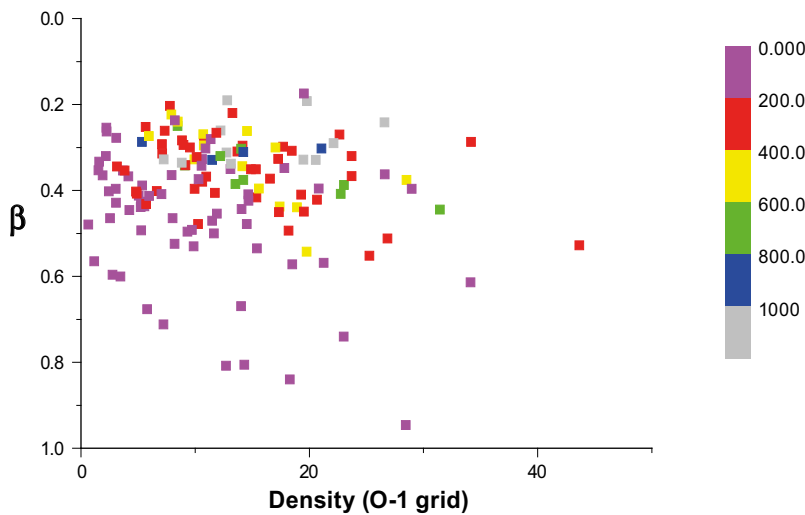

Figure 6-3. Scaling Exponent β (defined in Equation 4-12) versus estimated density. Dots colored according to the number of data by dataset.
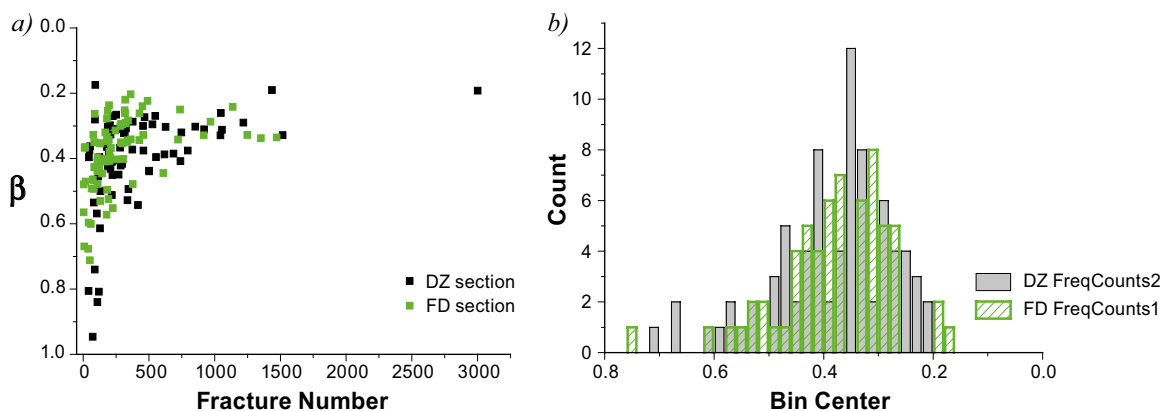

Figure 6-4. a) Scaling exponent β versus number of fractures in a dataset, colored according to DZ or FD. b) Histogram of β for the DZs (grey bars) and FDs (green bars) datasets.

Despite these datasets to datasets variations we next consider the scaling evolution of σ_{var} when averaged over several datasets. How this averaging process is done could still be debated. We presently simply consider the mean values, for each defined value of s, of $\sigma_{var}(s)$ normalized by $(1-s/L)$ and based on the contribution of all datasets having a total length smaller than s. This is applied to a selection of spatially close datasets including borehole KFM08C and its neighbors (see Table 6-2).

The averaged variations of $<\sigma_f(s)>/(1-s/L)$ and $<\sigma_d(s)>/(1-s/L)$ are plotted in Figure 6-5a and Figure 6-5b respectively. FDs and DZs datasets are treated separately. Both series display a well-defined mean behavior with an evolution with scale s consistent with a power-law scaling in $s^{-\beta}$ and β in the range [0.3; 0.35]. Despite a shift between FDs and DZs ($\sigma_{FD}(s)<\sigma_{DZ}(s)$), due to highest densities for the latter, their scale evolution is the same. Thus the variability scaling exponent is likely one characteristic of the natural fracturing process variability, which occurs independently from the intensity of fracturing (like intensity variation between FDs and DZs). Moreover, the natural spatial density variability of fracture systems appears significantly different – and larger – from the one associated to Poisson models.

Finally we analyze the possible dependence of β on the considered range of fracture orientations. β is computed simultaneously for entire datasets and for subsets based on the division into bins of fracture pole orientations. The division into subsets is based on the "O-3" grid defined in Figure 5-1b. Results are plotted in Figure 6-6, where the exponents β_i by orientation bin are plotted against the dataset reference exponent β. Dividing datasets into subsets obviously leads to a relative decrease of the amount of data into each subset, which automatically increases the apparent variations in these subsets. Despite this expected effect, there is a good correlation between variability scaling exponents per bin and the reference variability scaling exponent. The correlation is best observed for the sub-vertical bin, which is also the densest among the bins.

Table 6-2. Selection of datasets from KFM08C and its surrounding boreholes, including KFM08B, KFM08C and KFM07A.

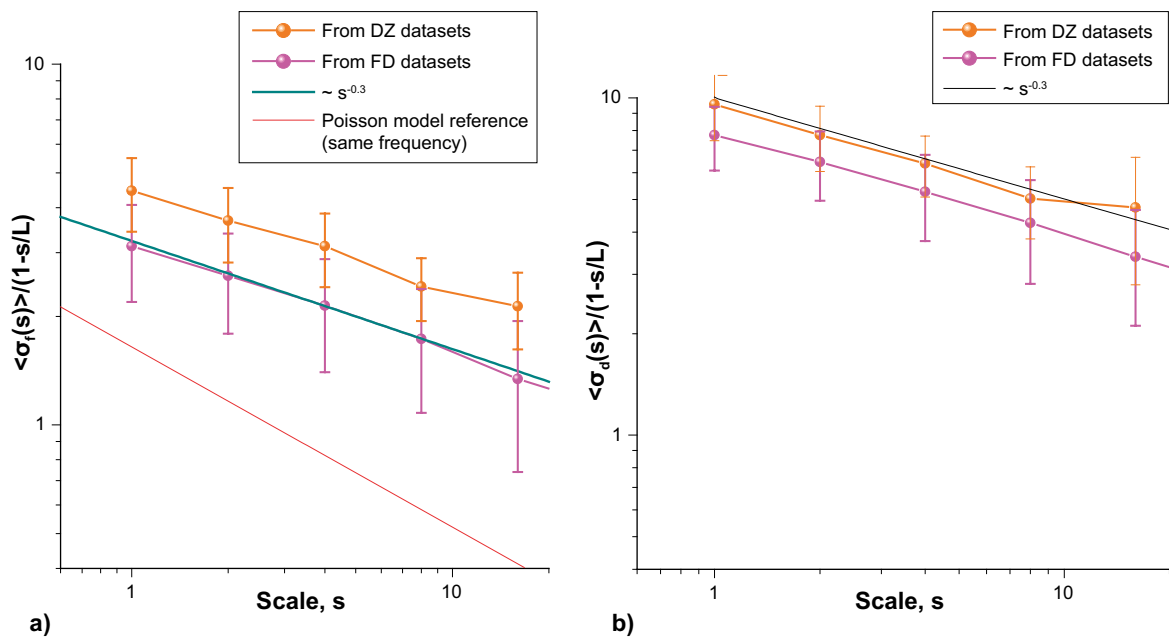| KFM07A-Dza	KFM07A-DZb	KFM07A-DZc	KFM07A-DZd	KFM07A-FFM01a	KFM07A-FFM01b
KFM08B-DZ1	KFM08A-DZ1	KFM08A-DZb	KFM08A-FFM01	KFM08B-FFM01	KFM08B-FFM02
KFM08C-DZ1	KFM08C-DZ2	KFM08C-DZ3	KFM08C-FFM01a	KFM08C-FFM01b	KFM08C-FFM01c
KFM08C-FFM01d	KFM08C-FFM06				

Figure 6-5. *Averaged variations of a) $<\sigma_f(s)>/(1-s/L)$ and b) $<\sigma_d(s)>/(1-s/L)$ over the datasets selection defined in Table 6-2.*
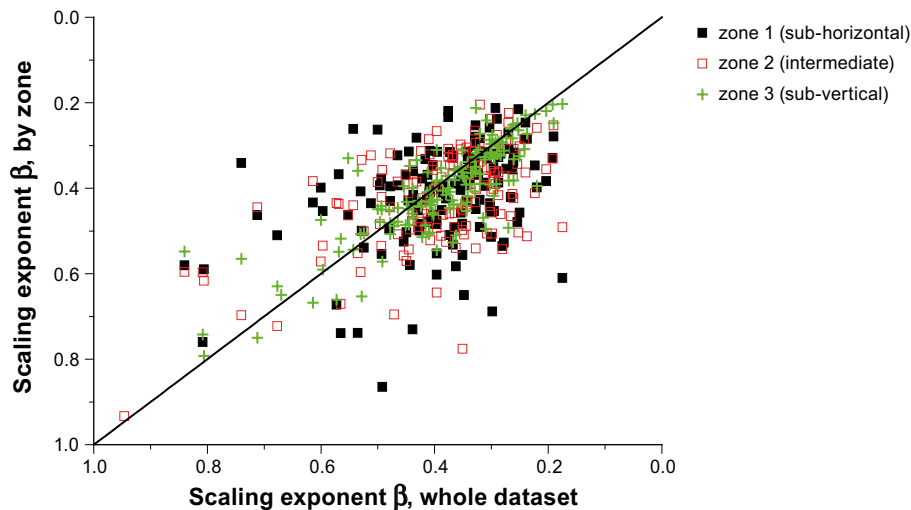
Figure 6-6. *Scaling exponents by orientation bin versus "whole density" scaling exponent. Orientations division into 3 bins from sub-horizontal to sub-vertical dips values, respectively with dips in bin 1: [0;25], bin 2: [25;65] and bin 3: [65;90].*

In this section, we have tried to identify, through a scale analysis, the specific signature of the fracture density spatial variability. Results are consistent for both DZs and FDs and for varying bins of orientations. Despite important fluctuations, which are anyway expected from a stochastic process like the Poisson model in similar conditions (i.e. datasets sizes), there are good indications toexpect a scaling of the density variability consistent with a power law of exponent β, on average equal to 0.3 to 0.35. This denotes a variability significantly larger than what would arise from a Poisson model and is likely the sign of intrinsic correlations between the fracturing properties. In other words, this property would be the signature of the real fracturing process. At this stage of investigations, there is no well-defined model to explain it, although preliminary work on weak fractal DFN models likely ends up into compatible observations.

The consequences of the fracture density spatial variations can nevertheless be quantified and integrated into the datasets density distribution estimates which are next combined into the classification process.

6.3   Grid of analysis and preliminary results

Pole clustering of fractures is commonly observed in field data, with most clusters close to horizontally or vertically oriented fractures. The intermediate bin between "close to" horizontal and "close to" vertical poles rarely contains a dense cluster of preferential orientations. Besides, when the amount of data is significant, it is common to observe some displacement (in term of orientations) of the apparent fracture preferential orientations from one observation location to the other, whereas these fracture clusters may come from unique episodes of a site geological and stress history. However, the link between fracture formation or geological history and "moving clusters" is difficult to unambiguously established and may become irrelevant in case of too big variations to be modeled into single DFN models. Also, things get worst when the amount of data and potential number of clusters increases.

In this context, our approach allows a systematic analysis of fracture densities versus orientations, from various locations and datasets, and eventually from various orientation grids of analysis. Once an orientation grid of analysis is defined, its bins boundaries are fixed and fracture densities are accordingly computed, depending on fracture pole orientations. However, even if the grid bins are fixed, the contribution of each fracture to the bins may be distributed through several bins when a distribution is associated to each fracture pole rather than a unique value (Darcel et al. 2009). In doing this, the bins boundaries become fuzzy.

A grid of analysis (Figure 5-1) can be defined without prior information or possibly be defined to reflect as best as possible the presumed geological footprint, or some modeling application specific objectives. Defining the "best" grid of analysis or the best division of the fracture density into orientation families is not trivial and may be unrealistic for usual conditions. The example of borehole KFM08C is used as an illustration.

Discrete and contoured stereonets of the KFM08C datasets are plotted in Figure 6-7 and Figure 6-8 respectively. This classical stereonet representation is incomplete: it does not include the total fracture density or the variability estimate. But it still highlights the occurrence and intensity of pole clusters relatively to each dataset taken individually. In the present case, we observe, on each section, apparent pole cluster maxima corresponding both to sub-horizontal or sub-vertical fractures. The sub-vertical pole maxima positions in the stereonet apparently differ from one section to the other.

The classification algorithm is applied for this set of datasets, first for the entire densities and then for an orientation grid including 5 bins (Figure 5-1d). This was presented in Darcel et al. (2012) and is recalled in Figure 6-9.
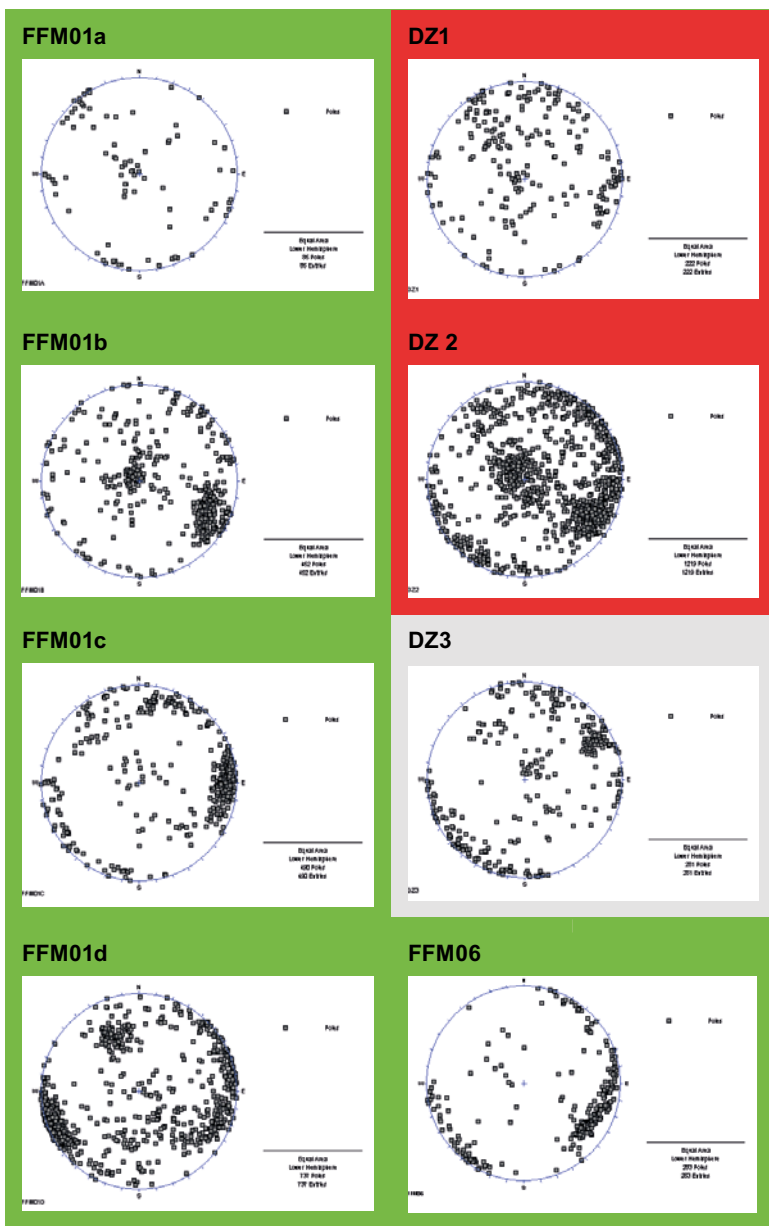
Figure 6-7. *Discrete stereonets of KFM08C sections. Colors refer to the three SFMs resulting from the classification process applied on the O-5 grid (Figure 5-1d).*
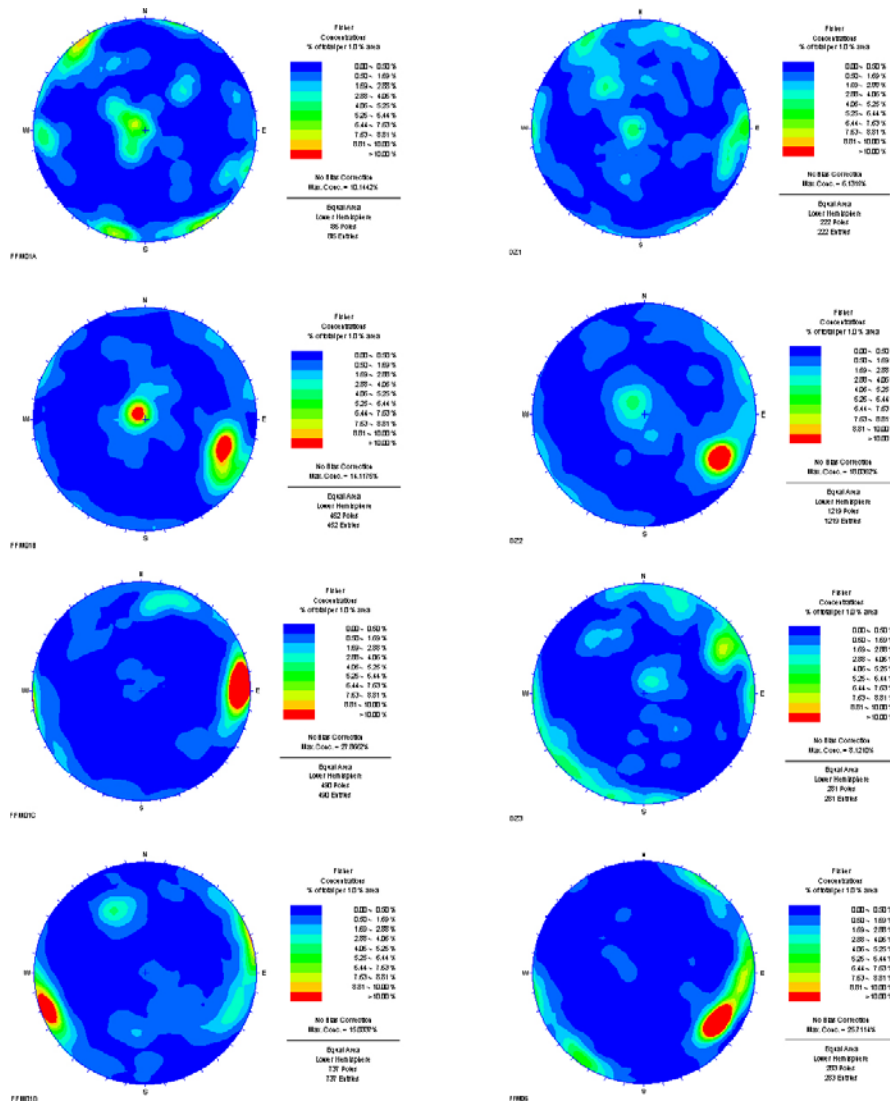
Figure 6-8. *Same as Figure 6-7, contoured stereonets; without orientation bias correction, but the bias is the same for each dataset since they come from a unique borehole.*
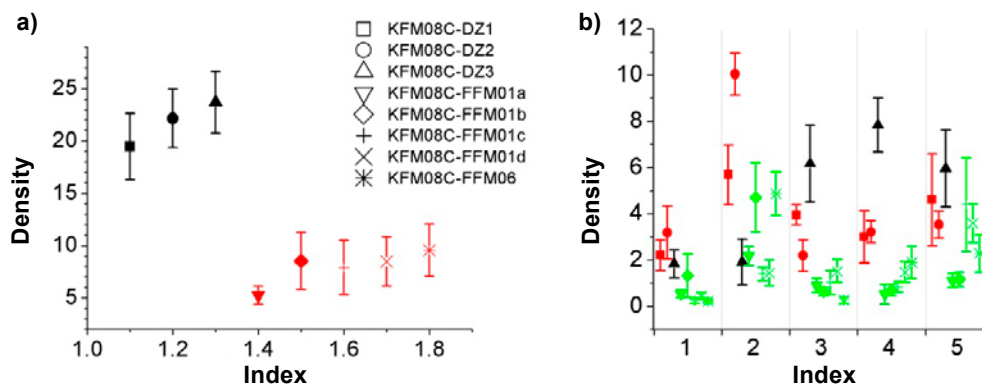
Figure 6-9. *Density distributions from Darcel et al. (2012):"KFM08C datasets density estimate and comparison in the SFD approach. Plots display density in y-axis and reference to bin number in x-axis; in each bin datasets values are evenly distributed for visualization purposes. Within each bin, each symbol represents a dataset, mean and standard deviation. Symbol colors indicate the class. a) Total density estimates: the 3 DZ datasets are grouped in one class of high density and the 5 FD datasets are grouped in a class of lower density. b) Same approach for a 5 bins study (Figure 5-1d). 3 classes arise: one isolated DZ, the two other DZs and finally all the FDs, which remain in the same class."*

The classification based on entire densities highlights the distinction between the DZs and FDs. This is not surprising since having a high density is one of the conditions to "be" a DZ. The second classification with 5 bins (Figure 5-1d) keeps the FD datasets into a single class (and SFM), whereas the 3 DZ datasets are divided into two classes (see illustration in Figure 6-7). When both density and variability are taken into account in the present case, with pole clusters potentially distributed into 5 cells, the FFM06 dataset of KFM08C is not statistically different from the other FFM01 datasets of KFM08C.

When the number of datasets increases and if the difference between datasets are less striking (for instance if only FFM datasets are considered, or only DZ datasets, the range of densities is likely diminished), it becomes unlikely to keep the "O-1" classification into the "O-5" one (Figure 6 and 7 in Darcel et al. 2012). Moreover, with a difference between two datasets distributions defined in Equation 5-1, i.e. defined as the average distance over N_z bins, increasing the number of bins leads to increase the possibility to compensate local differences from bin to bin. Applying the restrictive criterion (Equation 5-2) instead, with 5 bins, leads defining many classes, thus reflecting the high variability of properties.

Based on these preliminary observations, we limit the number of bins for applying the classification to all datasets at Forsmark. First the "O-1" is considered, and then the "O-3".

6.4   Preliminary results: classification based on entire density only

The "O-1" classification relies on comparisons between entire densities (mean and σ_{var} estimates) of datasets, without consideration of pole-clusters position variations. The initial 148 datasets are defined in Section 6.1. The corresponding evolution of Δ_{max} with the number of classes is displayed in Figure 6-10. At each step of the clustering process the two closest classes are merged and the new formed class is updated. For each class, one associates one dispersion value as the maximum distance between the datasets of the class and the class model itself. Therefore at each step of the clustering process the total number of classes is reduced by one and the maximum dispersion (Δ_{max}) is increased. Having a too large Δ_{max} – and too small number of classes- finally means that some classes have clustered too many – not compatible – datasets.

For Δ_{max}<0.5, clustering datasets together do not create a significant decrease of class cohesion (increase of Δ_{max}). From Δ_{max}>0.5 and above almost every new step of the clustering process does increase more significantly the distance within one class, leading to classes more and more dispersed. But there is no shift in the evolution of Δ_{max} such as the one observed from the tested Poissonian models (when 3 models are forced into 2 classes, Figure 5-4). We nevertheless use the Poissonian model reference to define a reasonable limit for Δ_{max} and stop the process when Δ_{max} is equal to 1.66, leading to 6 classes.
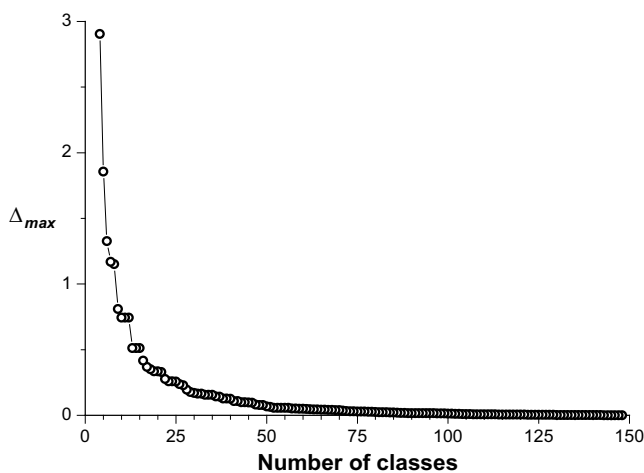

Figure 6-10. *Forsmark data. Evolution of* Δ_{max} *with the number of classes during the classification based on the O-1 grid.*

Once sorted according to the above limit, all the density distribution estimates are plotted in Figure 6-11. As expected from the soft evolution of Δ_{max}, there is no apparent density gap or jump between the identified groups of datasets. We rather observe a smooth variation of densities from the smallest to the largest densities. Classes 1 and 2, respectively having the largest and smallest densities, contain very few datasets. The four other main classes are logically ordered, with increasing density, from class 3 mostly made of FD datasets and up to class 5 dominated by DZ datasets. DZ datasets are in general denser than FD datasets, however many dense FD datasets display densities compatible with DZ, and vice versa.

In summary, in the present case, where only the entire densities are considered (i.e. without consideration of preferential orientations and clusters of fracture poles in the orientation distribution) we observe 6 classes whose respective limits are defined from the classification process but which are not identified by striking changes in estimated densities. There is rather a significant overlapping between distributions (the density estimate of one dataset may belong to the range of another dataset density estimate even if the latter belongs to a different class). This can be explained by the fact that at the scale of observation the probability is high of having internal variations larger that "external" ones (i.e. from one SFM to the other). In these conditions there is some overlapping between a class definition and the external but close datasets. In other words, the apparent limits between SFMs appear fuzzy. On the other hand, considering only entire densities may prevent from identifying SFM particularities and specific signature related to preferential orientations in the fracturing system. This is further investigated in the next section.

6.5   Description based on horizontal versus vertical fractures

The O-3 grid schemed in Figure 5-1b differentiates fractures according to their dip values, with 3 orientation bins whose limits are: dip equal to 25° and to 65°. Fractures are then sorted through 3 categories which correspond to "sub-horizontal", "intermediate" or "sub-vertical" trends.

In these conditions, the evolution of Δ_{max} with the number of classes is computed and results are plotted in Figure 6-12. As long as $\Delta_{max}<1$, the increase is regular and slow. Once $\Delta_{max}>1$ two relatively significant jumps of Δ_{max} are observed, for 7 and 9 classes (black and blue dots in Figure 6-12). The classification process is pursued until Δ_{max} is equal to 1.66 and the corresponding number of classes equal to 7 (black dot in Figure 6-12).

The SFM density distributions for the 7 classes are presented in Figure 6-13. The estimated mean and σ_{var} values are plotted for each bin of orientations. The composition of each SFM is summarized in Table 6-3 (and detailed in Appendix 2: Classification into SFMs). Within each of the 7 classes, the density of vertical fractures is larger than the one of sub-horizontal and intermediate orientations. The density of horizontal fractures is low in general but more significant for two classes, n°2 and n°4. Almost all classes – except the two smallest ones, classes 3 and 5 – combine both DZ and FD datasets (Table 6-3). However, in each case either the one or the other is dominant in the class. In particular, class 6 is made of DZ datasets up to 80% and class 7 is made of FD datasets up to 80%. Although less important, classes 1 and 2 are dominated by DZs while classes 3, 4 and 5 are dominated by FDs. The role of depth in the classification is further reviewed in the next section.

Table 6-3. Summary table for the 7 classes and SFM models from the Forsmark data based on the O-3 grid with Δ_{max}=1.66. DZ is for Deformation Zones (SKB definition) and FD is for Fracture Domain (SKB definition).

Class	Number of datasets in the class	Color in Figure 6-13 and Figure 6-16	Proportion of DZ datasets (%)	Comments
1	9	black	78	mostly DZ datasets and largest densities
2	17	red	65	DZ dominated and shallow depths
6	42	pink	81	DZ dominated
3	4	green	0	FD datasets dominated but few datasets
4	17	blue	29	FD dominated, shallow depths
7	51	brown	20	FD dominated
5	7	cyan	0	FD dominated, very low densities
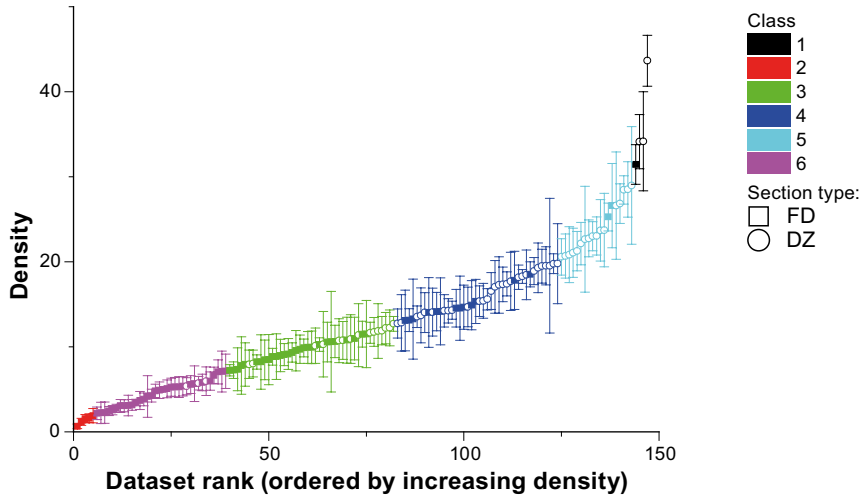
Figure 6-11. *Datasets densities (ordered by increasing density). Datasets densities colored by the class index obtained with the O-1 grid. DZ datasets as empty disks and FD datasets as filled squares.*
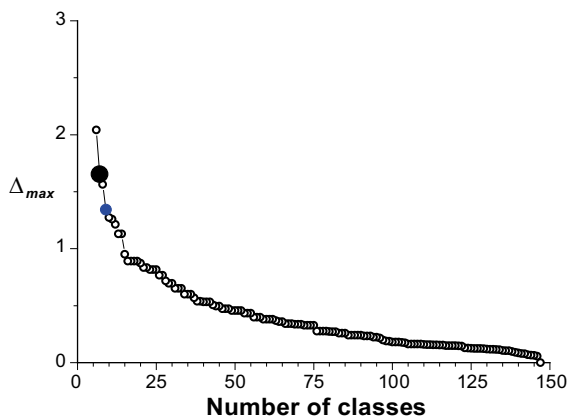
Figure 6-12. *Evolution of Δ_{max} with the number of classes during the classification based on the O-3 grid (recalled in Figure 5-1b).*
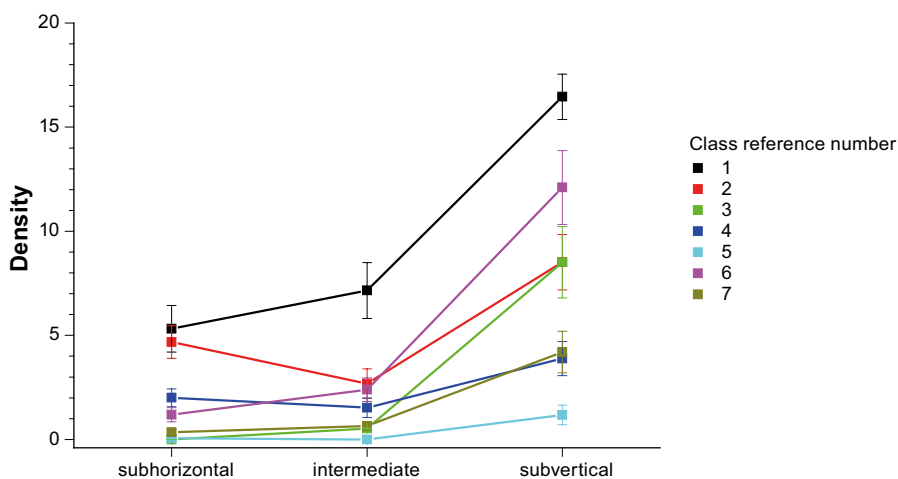
Figure 6-13. *7 SFMs of the Forsmark data based on the O-3 grid (Figure 5-1b) with Δ_{max}=1.66. See text for details.*

6.6 Densities and depth dependency

Density distributions of all the considered datasets, expressed as mean and σ_{var} values for each orientation bin, are plotted as function of dataset mean elevation in Figure 6-14a, -b and -c, respectively for the sub-horizontal, intermediate and sub-vertical bins. Results of the above-mentioned classification are highlighted by the colors chosen to reflect the belonging to common classes. This representation thus highlights the potential density evolution with depth of the 3 grid bins (sub-horizontal, intermediate and sub-vertical fractures).

The global evolution with depth of the fracture density – all datasets considered together and according to the vertical position of each one – is computed by orientation bin and plotted as a thick black line in the figure (Figure 6-14a, b and c). We thus observe a global decrease of the fracture density of horizontal fractures toward large depths (Figure 6-14a). From surface down to approximately 120 m, the horizontal fracturing density follows an exponential decrease whose characteristic scale is 60 m. Below this depth, the decrease is less important but still exponential now with a characteristic size equal to 500 m. There is no such depth effect observed for the two other bins of orientations, the intermediate and the sub-vertical ones.

The final classification into 7 SFMs reflects partly the depth dependency: in addition to the general picture, where classes are organized from large densities and mainly DZs to smaller densities and mainly FDs, two classes are identified for large densities and shallow depths (classes 4 and 2). Indeed, both class 4 and 7 are dominated by FDs, but class 4 is located at shallow depths and the density of horizontal fractures is larger. The same combination occurs for class 2 and 6, with class 2 having more horizontal fractures and shallower depths.

In the next sections (6.7 and 6.9) we review first the classification robustness with regards to the last steps of the classification (i.e. what if the classification were stopped at a higher number of classes) and then consider the effect of changing the orientation grid of observation.

6.7 Last stages of the classification, from 9 to 7 classes

Here we illustrate how the classification evolves at the last stages of the classification process. The intermediate case with 9 classes is compared to the final 7 classes' stage. Both stages correspond to a relatively important jump of Δ_{max}. When 9 classes are defined Δ_{max}=1.35, whereas it is equal to 1.66 for 7 classes.

The 9 classes classification is displayed in Figure 6-15 (one color by class), where horizontal fractures densities are plotted against depth. This figure can be directly compared to Figure 6-14a. Beyond a few changes for some datasets at classes boundaries, the main differences occur for class 6 and class 7: both these large classes are split into 2 sub-classes (6-1, 6-2, 7-1, and 7-2). Class 7-2 is mainly composed by datasets from FFM01 (80%) while 7-1 is composed mainly by datasets belonging to other FFM (80%). Therefore, although fracturing properties of most FFM are very close (and largely merged into one to two big classes in the final description), the higher refinement allows differentiating further the properties. The class 6, dominated by DZ datasets, is divided into two classes named 6-1 and 6-2. Class 6-1 is close to class 4 (FD shallow with a relative high density of horizontal fractures) while class 6-2 is closer to class 7 (other FD datasets). These two classes are in the further steps merged together because of their similar densities for both intermediate and vertical fractures.

Finally, the classification can be considered as robust: there are few variations between the main SFMs at the latest stages of the classification. Nevertheless, defining which final Δ_{max} provides the best optimization is not trivial.

Figure 6-14. *Evolution of fracture density with depth with the O-3 classification. The 7 final classes are represented by the different colors (see inlet), density vs. depth of a) sub-horizontal fractures, b) intermediate fractures and c) sub-vertical fractures. The "depth error bars" represent the borehole section min and max depth.*
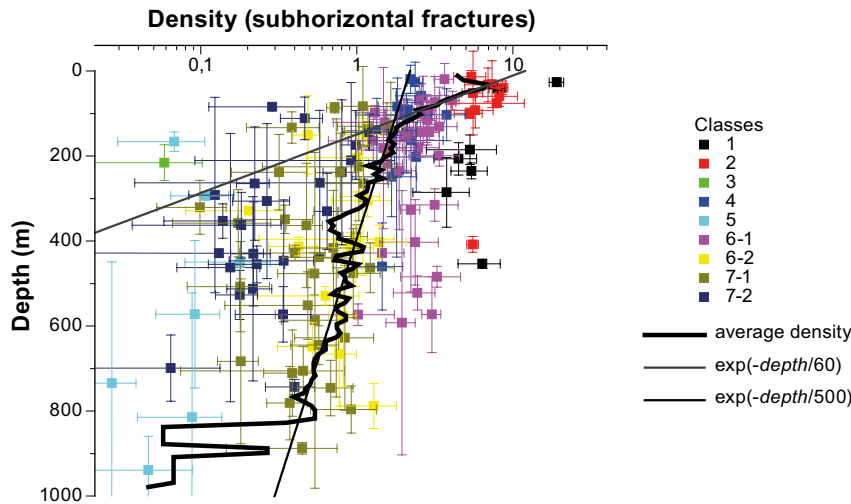
Figure 6-15. *Evolution of fracture density with depth for the subhorizontal fractures. Datasets are colored according to the O-3 classification, but with Δ_{max}=1.35.*

6.8 Spatial organization of the SFMs

The spatial distribution of SFMs is displayed in Figure 6-16. Density values are plotted on three 2D maps, each representing a certain range of elevation values: from surface down to 118 m, from 118 m down to 592 m and finally below 592 m. For each layer, density values are plotted as "intensity bullets" whose center is located at the corresponding dataset Northing/Easting position. Each bullet is colored by class index and its size is proportional to density (of horizontal fractures for the 0–118 m layer, vertical fractures for the two other layers).

We observe that the surface effect (horizontal fracturing density increase, classes 2 and 4) is mainly localized in the south-western part of the area of interest. At larger depth (below 118 m) the fracturing properties are dominated by a mix of classes 7 *(FD dominated)* and 6 *(DZ dominated)*. The DZ do not seem to act as domain boundaries. Finally, at the deepest elevations (Figure 6-16c) densities are rather low, with few differences between FD and DZ datasets.

6.9 Alternate description based on one horizontal and two vertical directions

The previous analyses (sections 6.5 to 6.7), based on dip values only, have not considered the role of strikes in the definition of the SFMs. We finally modify the O-3 grid and replace the former "vertical" and "intermediate" grid bins by two vertical grid bins (name "O-3hv"). The horizontal bin is unchanged. The previous intermediate fracture bin is merged into the two new vertical bins. The two new vertical bins are defined by two major strike directions (Figure 6-17c). The limits of the vertical bins are chosen according to dominant strike direction of pole clusters from the entire site (i.e. all poles of all boreholes, Figure 6-17a and b).

The classification process is applied until Δ_{max}>1.66, which yields 7 classes whose densities are plotted in Figure 6-18. Density distributions of all the considered datasets, expressed as mean and σ_{var} values for each (horizontal, vertical bin 2, vertical bin 3) orientation bin, are plotted as function of dataset mean elevation in Figure 6-19a, -b and -c (similar to Figure 6-14; but only the horizontal bin is unchanged between the O-3 and the O-3hv cases). Results of the O-3hv classification are highlighted by the colors chosen to reflect the belonging to common classes, the color code is the same in Figure 6-18 and Figure 6-19.

a) Depth 0–118 m
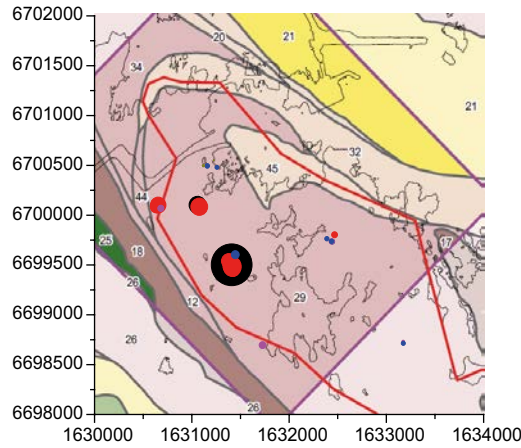
b) Depth 118–592 m
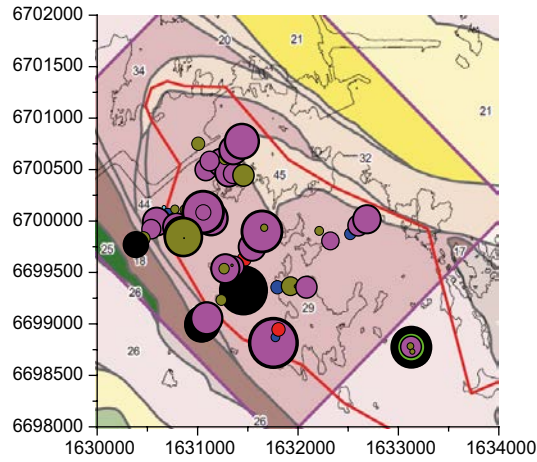
c) Depth 592–1,000 m
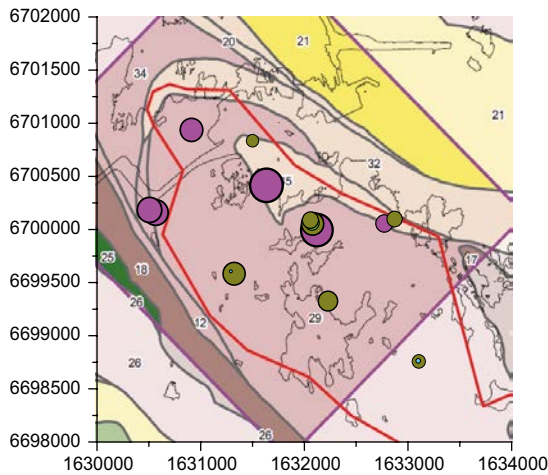
Figure 6-16. *Dataset estimated densities on three 2D maps, each representing a certain range of elevation values: a) density of "horizontal" fractures, occurring from surface down to 118 m, b) density of vertical fractures occurring from 118 m down to 592 m and finally c) density of vertical fracture occurring below 592 m. See text for more details. The background map comes from Olofsson et al. (2007) and the area of interest is represented by the red line.*
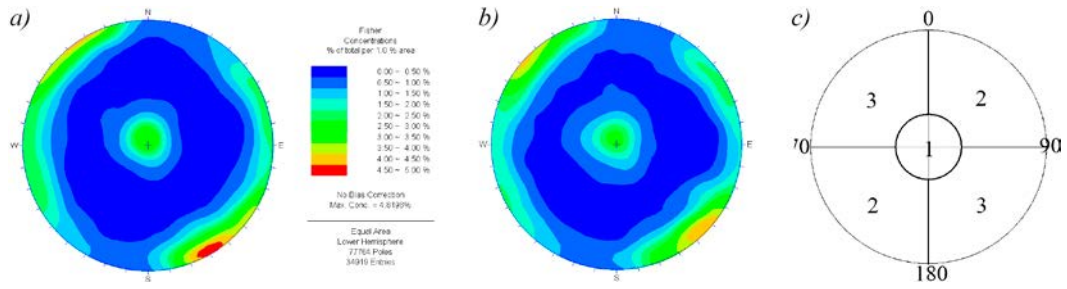
Figure 6-17. *Contoured orientations in Forsmark for fractures from a) FDs and b) DZs. c) The alternate O-3 grid of analysis where orientation bins are named as grid zone 1, 2 and 3.*
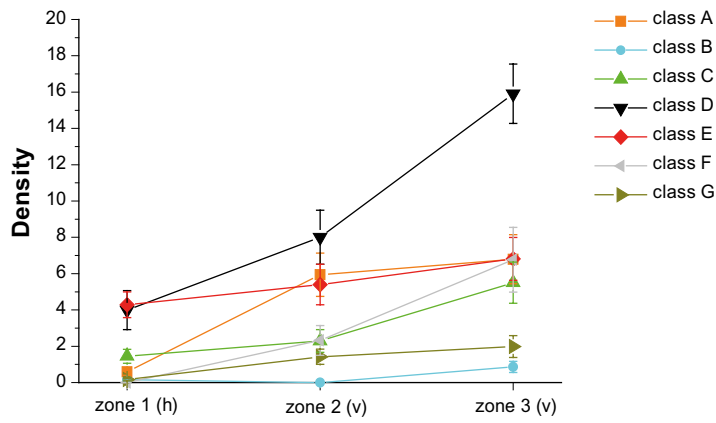
Figure 6-18. *Classes density distributions from the O-3hv classification, 7 classes defined. See text for more details.*
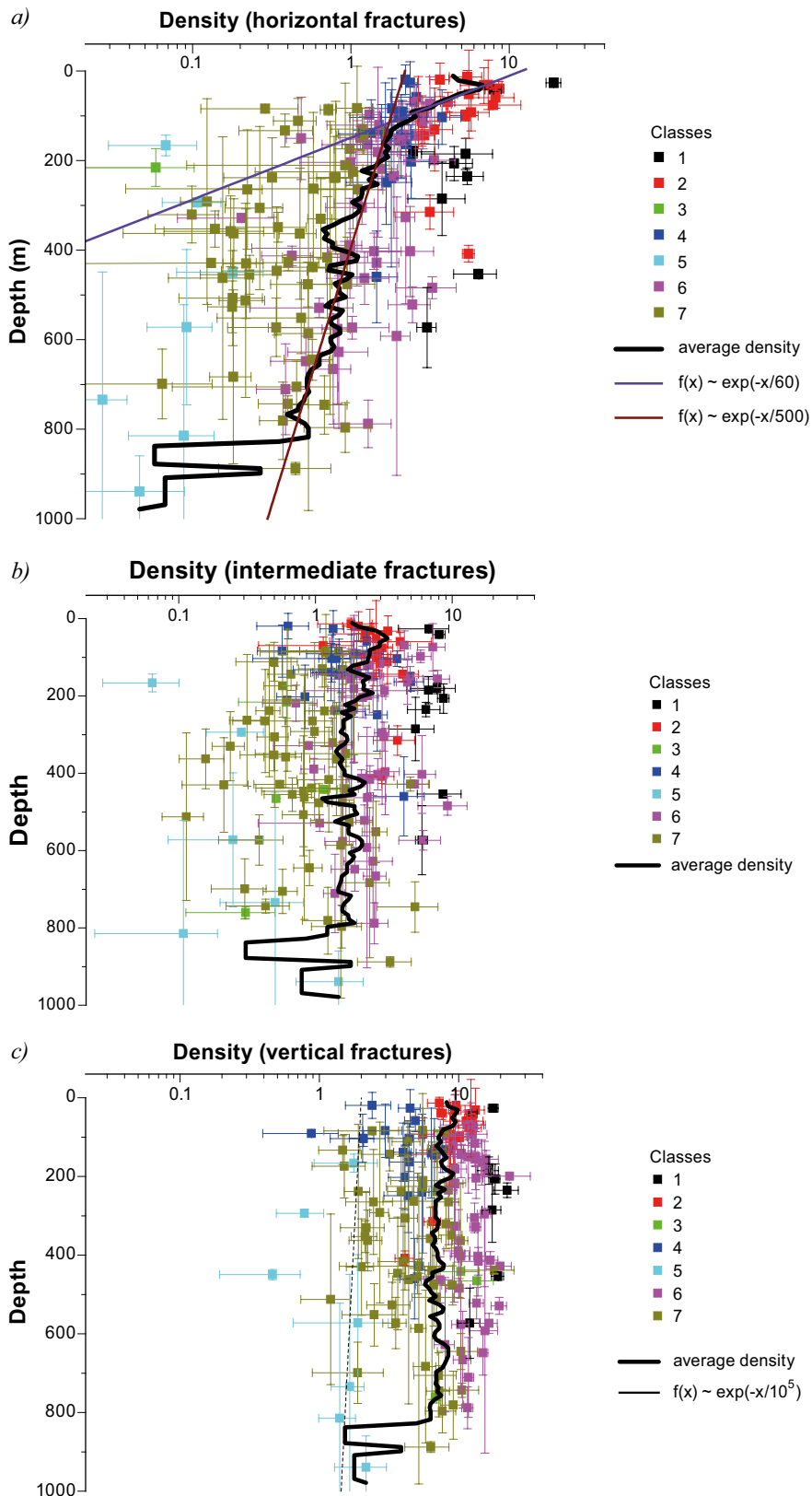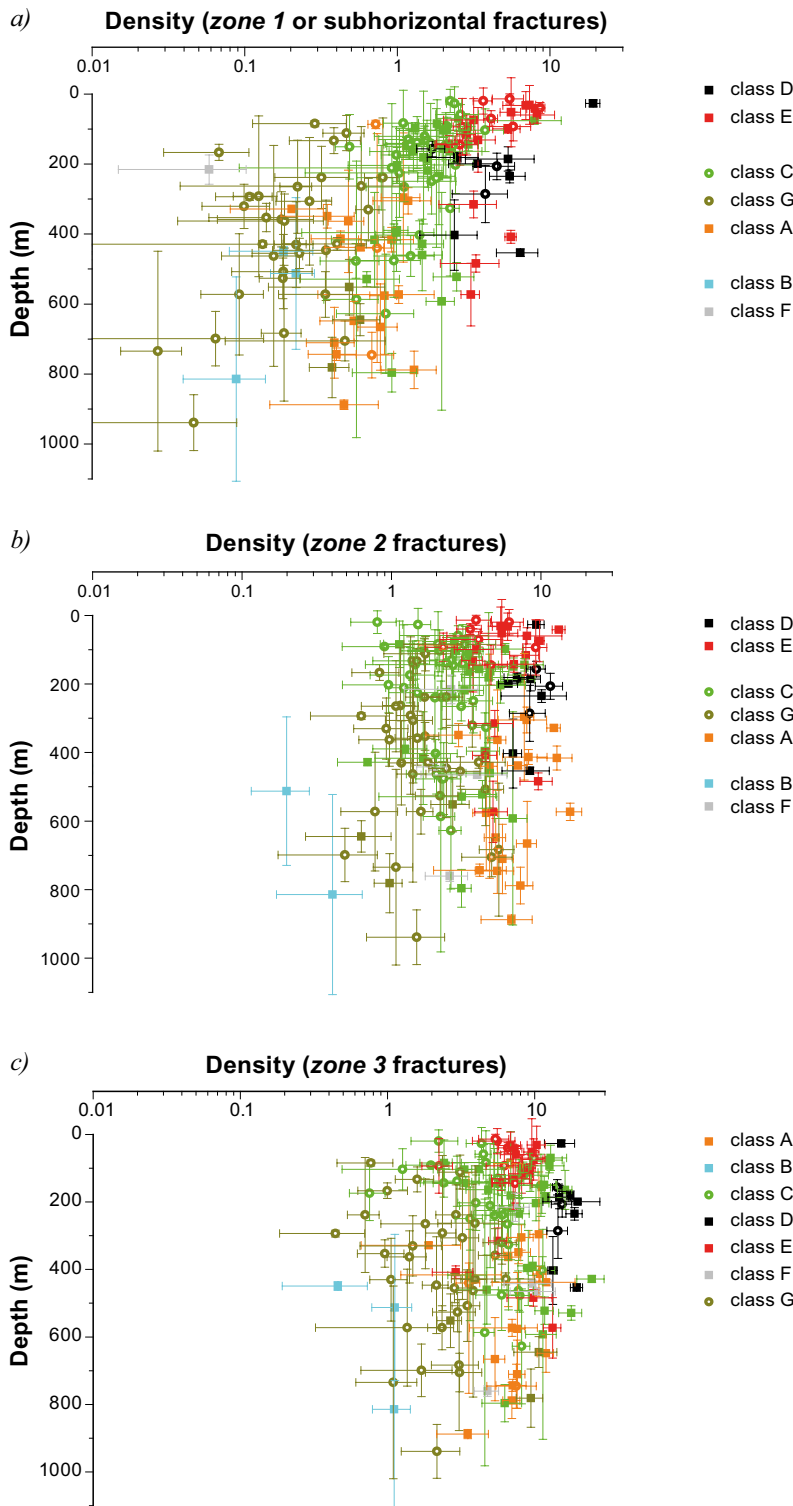
Figure 6-19. *Evolution of fracture density with depth with the O-3hv classification. The 7 final classes represented by the different colors (see inlet), density vs. depth of a) sub-horizontal, b) vertical – bin 2 and c) bin 3 fractures. The "depth error bars" represent the borehole section min and max depth. Grid bins of the alternate O-3 grid are recalled in Figure 6-17.*

We observe that the former class 1, with very high densities of DZ and the former class 2, with mainly DZ datasets displaying the horizontal fracturing surface effect, as recalled in Figure 6-13 and Table 6-3, are still found in the present classes D and E respectively.

With some limited recombinations, the former class 7 (and even class 7-2, i.e. low densities of 7) is now associated to former class 5 to define the new class G. This class G is mainly composed of "FFM01" datasets. Also a large part of former class 6 (even class 6-2, so the lower densities of 6) is now transformed into class A, largely characterized by a low density of horizontal fractures.

The most important difference in the classification is reflected by the new class C: this one is now a mix between former class 7-1, former class 6-1 and former class 4. This class thus is composed in equal parts by FD and DZ datasets (50%–50%). It is mainly characterized by a preponderant density for bin 3 (i.e. vertical fractures striking NE).

The consideration of several distinct classification schemes improves the understanding of the fracture site organization. The O-3 scheme was adapted to clearly identify a surface effect like the density increase of horizontal fracturing close to surface. The influence of the surface effect in the final classification is nevertheless less visible in the second classification where 2 of the 3 orientation bins are for vertical fractures, thus counterbalancing the weight of the horizontal one. In this latter case, the high density of vertical fractures in bin 3 (i.e. vertical fractures striking NE) becomes decisive to define class C.

7 Summary and conclusions

The spatial variability of fracture density is a key issue for the DFN modeling. This project aimed to review and upgrade the site DFN modeling approach initially introduced in Darcel et al. (2009) especially with regards to relations between models, model assessment from data and spatial variability of properties. In this perspective the first stage was dedicated to DFN models and modeling methodology. In the second stage real data, from Forsmark site, were investigated to initially appraise the intrinsic spatial variability and scaling behavior of fracture properties. Then the classification approach and the class SFMs properties assessment were applied to a DFN analysis of the Forsmark site.

Poisson DFN models are commonly used to mimic the spatial variation of fracture density. For such cases, we derive the relation between the fracture density estimate standard deviation σ_{var} and the scale of observation: on average σ_{var} evolves like the square root inverse of the scale of observation. Thus, the spatial variability of Poisson DFN models is quantified and one can used it as a reference case. The other considered DFN models differ from the Poisson one through the addition of a fractal dimension assigned to the spatial distribution of fracture positions (Darcel et al. 2003a). Thus defined, the fractal DFN models display an apparent slower decrease of σ_{var} with scale and then lead to larger standard deviations of the density estimate distribution at constant number of data into a dataset. At this point, further investigations are required to fully define the evolution of σ_{var} with scale and fractal dimension. In both Poisson and fractal cases, increasing the number of elements into a datasets, either by sampling a higher density zone, or by increasing the section of observation length, obviously improves the accuracy of the density estimate (and reduces σ_{var}).

Analyses of real datasets from Forsmark display a wide range of density distribution estimates. But in many cases, they are consistent with scaling of σ_{var} in scale size to the power –0.3, which is clearly different from Poisson models and closer to weakly fractal DFN model. We consider that the found variability scaling exponent is likely the signature of real fracture properties at Forsmark.

The complete approach is extensively used for synthetic references cases and real data. It includes first assessment of density estimate distribution for each dataset, then classification into groups of similar datasets from which SFMs are derived and finally possible changes in the analysis controlling parameters (dataset division into a number of ranges of fracture orientations – the orientation bins of the grid of analysis, Δ_{max}).

From the real conditions analyses, with here close to 150 initial datasets and more than 60,000 data, it appears that the limits between SFMs can be fuzzy or equivalently, the final SFMs partly overlapping each other. As a consequence, small changes in the parameters may moderately change the SFMs limits and thus there is not a unique final classification and subsequent set of classes and SFMs. This is the straight consequence of real conditions, where variations of fracture density cover a wide range of values and spatial variations can be important both at small and larger scales. These conditions leads to use a limited number of orientation bins to define each density distribution, so that each density estimate still relies on a large number of data and therefore rather well defined estimates.

Once this is stated, we recall that there are limits to the DFN modeling refinement: because of properties spatial variations, the access and knowledge of underlying models necessarily is limited. This is all the point of the method. At final stages of the classification process, each new class merging step induces larger and larger Δ_{max} jumps. This reflects that each time bigger and bigger classes are merged and internal dispersion within a class increased. Changing the final threshold value of Δ_{max} or even changing the orientation grid of analysis ends up into some changes in the final classification, which are limited and do not put into question the main picture. We recommend that, given the complexity of real in situ fracture organization, site scale analyses should be done from several classifications attempts (definition of the orientation cells in the grid of analysis and final stages of the clustering into classes). In these conditions, the method has further proven its efficiency to make objective analyses from any number of initial data, while properly including the spatial variability on the density estimates.

We finally summarize the principal outcomes of the Forsmark site analysis. Five main SFMs are derived. Few key elements control the general picture. Firstly, dataset total density is of major importance. As a consequence, highest densities SFMs are dominated by DZs whereas lowest ones are made of FDs. However, in the wide space between highest and lowest densities, SFMs contains both DZs and FDs, which illustrates that there is not always a clear distinction between both, from the DFN statistical representation view. Second, there is a clear depth dependency of the density of horizontal fractures: shallow zones exhibits a larger density of horizontal fractures, in particular in the south-western part of the area (class 2 and class 4 from Table 6-3). Two characteristic depths are thus identified, at 60 m and 500 m.

8   Perspectives

The work presented here has proven the method efficiency to encompass a modeling of spatial variability both within SFMs and naturally in-between SFMs. It has also highlighted some continuing issues:

- The variability of fracture density is likely reflected by a specific signature and evolution as observation size to the power –0.3. The confidence in such a characteristic could be better evaluated through analyses of tightened sampling conditions.

- Poisson DFN models underestimate the real local variations of fracturing properties. Fractal-like model are likely better adapted to reflect reality, however only preliminary analyses have been performed. This work should be pursued.

- The division of a site into spatial Fracture Domains each having their properties cannot reproduce a site scale trend evolution of density like the evolution with depth observed. How to combine such a site scale trend to Fracture Domains and or SFMs identified has to be developed.

- Local spatial variations of fracture densities defined from specific ranges of pole orientations were investigated here. However, the possible spatial correlations between the orientation of a fracture and its neighbors were not investigated. This should be considered into a revision of the spatial correlation model of SFMs.

9 Abbreviations

DFN   Discrete Fracture Network

SFM   Statistical Fracture Model

FD   Fracture Domain (Olofsson et al. 2007)

DZ   Deformation Zone (Olofsson et al. 2007)

sdev   Standard deviation

O-n   Orientation grid based on n bins

3D   Three Dimensional

References

SKB's (Svensk Kärnbränslehantering AB) publications can be found at www.skb.se/publications.

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

Chelcea S, Bertrand P, Trousse B, 2006. A new agglomerative 2–3 hierarchical clustering algorithm. In Baier D, Wernecke K-D (eds). Innovations in classification, data science, and information systems: proceedings of the 27th annual conference of the Gesellschaft für Klassifikation e.V., Brandenburg University of Technology, Cottbus, 12–14 March 2003. Berlin: Springer, 3–10.

Darcel C, Bour O, Davy P, 2003a. Stereological analysis of fractal fracture networks. Journal of Geophysical Research: Solid Earth 108. doi:10.1029/2002JB002091

Darcel C, Bour O, Davy P, de Dreuzy J R, 2003b. Connectivity properties of two-dimensional fracture networks with stochastic fractal correlation. Water Resources Research 39. doi:10.1029/2002WR001628

Darcel C, Davy P, Bour O, de Dreuzy J-R, 2004. Alternative DFN model based on initial site investigations at Simpevarp. SKB R-04-76, Svensk Kärnbränslehantering AB.

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, Le Goc R, de Dreuzy J-R, Bour O, 2009. Statistical methodology for discrete fracture model – including fracture size, orientation uncertainty together with intensity uncertainty and variability. SKB R-09-38, Svensk Kärnbränslehantering AB.

Darcel C, Davy P, Le Goc R, 2012. Statistical Fracture Domain methodology for DFN modeling applied to site characterization. In Proceedings of Eurock 2012, the ISRM International Symposium, Stockholm, Sweden, 28–30 May 2012.

Haight F A, 1967. Handbook of the Poisson distribution. New York: Wiley.

Laplace P S de, 1814. Théorie analytique des probabilités. 3rd ed. Paris: Courcier.

La Pointe P, Fox A, Hermanson J, Öhman J, 2008. Geological discrete fracture network model for the Laxemar site. Site descriptive modelling, SDM-Site Laxemar. SKB R-08-55, Svensk Kärnbränslehantering AB.

Moivre A de, 1756. The doctrine of chances, or, A method of calculating the probabilities of events in play. London: Printed for A. Millar.

Olofsson I, Simeonov A, Stephens M, Follin S, Nilsson A-C, Röshoff K, Lindberg U, Lanaro F, Fredriksson A, Persson L, 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 inter-disciplinary coordination. SKB R-07-15, Svensk Kärnbränslehantering AB.

Wahlgren C-H, Curtis P, Hermanson J, Forssberg O, Johan Ö, Fox A, La Pointe P, Drake H, Triumf C-A, Mattsson H, Thunehed H, Juhlin C, 2008. Geology Laxemar. Site descriptive modelling, SDM-Site Laxemar. SKB R-08-54, Svensk Kärnbränslehantering AB.

Forsmark datasets

IDCODE	number	height	min_pos	max_pos
KFM01A_FFM01a	112	90.03	290.49	380.52
KFM01A_FFM01b	140	219.96	410.10	630.06
KFM01A_FFM01c	185	300.42	678.69	979.11
KFM01A_FFM02a	397	93.18	102.98	196.15
KFM01A_FFM02b	303	92.87	196.83	289.70
KFM01A_ZFMENE1192	122	28.21	381.40	409.61
KFM01A_ZFMENE2254	256	46.67	631.35	678.02
KFM01B_FFM01a	346	190.54	223.79	414.32
KFM01B_FFM01b	118	45.58	452.22	497.80
KFM01B_FFM02a	459	82.53	57.17	139.70
KFM01B_FFM02b	136	83.27	140.01	223.28
KFM01B_ZFMA2	434	40.59	16.23	56.83
KFM01B_ZFMNNW0404	251	36.86	415.04	451.90
KFM01C_8ZFMA2	399	31.32	66.63	97.95
KFM01C_FFM01a	841	51.14	252.19	303.33
KFM01C_FFM01b	1574	112.58	329.00	441.58
KFM01C_FFM02a	174	19.27	12.28	31.55
KFM01C_FFM02b	245	21.14	45.47	66.62
KFM01C_FFM02c	462	54.79	98.49	153.28
KFM01C_FFM02d	580	77.75	156.06	233.81
KFM01C_ZFMENE0060A	265	17.92	234.08	251.99
KFM01C_ZFMENE0060C	508	25.30	303.54	328.84
KFM01C_ZFMENE1192	364	12.94	32.33	45.27
KFM01D_FFM01a	375	217.16	191.13	408.30
KFM01D_FFM01b	60	53.00	420.03	473.03
KFM01D_FFM01c	364	193.43	475.15	668.58
KFM01D_FFM01d	160	91.11	699.00	790.11
KFM01D_FFM02	311	74.80	92.35	167.15
KFM01D_PDZa	157	22.44	168.65	191.09
KFM01D_PDZb	58	11.08	408.85	419.94
KFM01D_ZFMENE0061	144	29.87	668.74	698.60
KFM02A_DZ2	102	20.22	101.53	121.75
KFM02A_DZ3	174	23.98	159.50	183.48
KFM02A_DZ6	646	102.88	417.35	520.22
KFM02A_DZ7	260	79.86	521.19	601.05
KFM02A_DZ8	57	13.77	893.26	907.03
KFM02A_FFM01a	249	288.25	601.36	889.61
KFM02A_FFM01b	145	80.66	911.72	992.37
KFM02A_FFM03a	80	37.12	122.21	159.33
KFM02A_FFM03b	471	230.68	184.80	415.48
KFM02B_DZ1	120	16.32	98.53	114.85
KFM02B_DZ2	481	58.42	145.02	203.44
KFM02B_DZ3	157	18.65	412.10	430.75
KFM02B_DZ5	170	10.75	462.26	473.01
KFM02B_DZ6	308	24.93	486.50	511.43
KFM02B_RU1a	14	8.96	88.95	97.91
KFM02B_RU1b	156	29.38	115.36	144.74
KFM02B_RU1c	175	90.19	204.11	294.30
KFM02B_RU1d	5	10.05	300.31	310.35
KFM02B_RU1e	88	18.15	432.76	450.91
KFM02B_RU1f	22	9.69	451.05	460.74

IDCODE	number	height	min_pos	max_pos
KFM02B_RU1g	12	8.70	475.30	484.00
KFM02B_RU1h	108	58.87	512.57	571.45
KFM02B_RU2a	161	80.18	330.68	410.86
KFM04A_FFM01	1459	522.17	461.71	983.88
KFM04A_FFM04	227	25.63	174.84	200.46
KFM04A_ZFMA2	472	39.88	200.90	240.77
KFM04A_ZFMNE1188	1637	218.71	241.63	460.33
KFM04A_ZFMNW1200	525	65.07	109.66	174.73
KFM05A_DZ2	192	40.34	395.46	435.80
KFM05A_DZ3	1098	205.64	590.22	795.86
KFM05A_DZ4	543	107.03	892.50	999.53
KFM05A_FFM01a	208	158.66	235.71	394.37
KFM05A_FFM01b	169	149.25	439.71	588.95
KFM05A_FFM01c	88	96.03	796.16	892.19
KFM05A_FFM02	521	132.20	102.95	235.15
KFM06A_DZ1	224	43.34	102.28	145.62
KFM06A_DZ10	102	22.66	882.13	904.78
KFM06A_DZ11	321	70.88	925.38	996.26
KFM06A_DZ2	1067	162.17	195.81	357.98
KFM06A_DZ7	542	69.66	740.18	809.85
KFM06A_FFM01a	123	48.65	146.03	194.68
KFM06A_FFM01b	1023	381.19	358.75	739.94
KFM06A_FFM06a	207	71.56	810.26	881.82
KFM06A_FFM06b	56	19.74	905.12	924.86
KFM06B_DZ1	340	37.40	55.14	92.54
KFM06B_FFM02a	208	47.86	7.10	54.97
KFM06B_FFM02b	11	2.95	94.16	97.11
KFM06C_DZ1	393	66.55	102.19	168.75
KFM06C_DZ3	872	143.35	413.59	556.94
KFM06C_DZ5	423	53.53	623.28	676.80
KFM06C_FFM01	937	242.90	169.40	412.30
KFM06C_FFM05	454	90.50	906.28	996.78
KFM06C_FFM06a	246	65.16	557.74	622.90
KFM06C_FFM06b	1092	228.56	677.04	905.60
KFM07A-Dza	705	107.05	102.04	209.09
KFM07A-DZb	529	163.53	411.50	575.02
KFM07A-DZc	769	129.52	792.44	921.96
KFM07A-DZd	615	71.70	922.07	993.77
KFM07A-FFM01a	347	200.84	210.04	410.88
KFM07A-FFM01b	208	213.73	576.98	790.71
KFM07B-Dza	82	6.13	51.53	57.66
KFM07B-Dzb	117	8.63	93.19	101.82
KFM07B-Dzc	164	15.43	119.18	134.61
KFM07B-Dzd	161	17.96	226.51	244.47
KFM07B-FFM01a	29	28.33	196.65	224.98
KFM07B-FFM01b	248	51.94	245.38	297.32
KFM07B-FFM02a	374	45.60	5.39	50.99
KFM07B-FFM02b	182	34.81	58.04	92.85
KFM07B-FFM02c	85	15.23	102.95	118.18
KFM07B-FFM02d	252	59.08	135.45	194.53
KFM07C-Dza	606	79.86	308.09	387.95
KFM07C-Dzb	78	9.65	429.15	438.81
KFM07C-FFM01a	458	184.55	123.09	307.64
KFM07C-FFM01b	210	40.70	388.02	428.72
KFM07C-FFM01c	256	58.69	439.42	498.11
KFM08A-DZ1	1514	228.50	103.36	331.86

IDCODE	number	height	min_pos	max_pos
KFM08A-DZb	1128	185.28	764.39	949.67
KFM08A-FFM01	1623	431.00	332.19	763.19
KFM08B-DZ1	192	28.05	156.56	184.61
KFM08B-FFM01	406	110.90	45.03	155.93
KFM08B-FFM02	123	38.85	6.12	44.97
KFM08C-DZ1	239	29.44	161.13	190.56
KFM08C-DZ2	1307	122.13	419.31	541.43
KFM08C-DZ3	360	31.57	673.25	704.82
KFM08C-FFM01a	92	58.44	102.48	160.93
KFM08C-FFM01b	535	151.01	191.08	342.09
KFM08C-FFM01c	509	130.57	542.41	672.97
KFM08C-FFM01d	850	245.24	705.05	950.29
KFM08C-FFM06	304	76.20	342.45	418.65
KFM08D-Bsa	235	83.62	64.33	147.96
KFM08D-BSb	771	95.88	148.01	243.89
KFM08D-BSc	333	57.01	244.43	301.44
KFM08D-Bse	988	181.93	360.05	541.98
KFM08D-BSf	4229	399.74	542.01	941.75
KFM09A_FFM01a	331	88.45	124.35	212.80
KFM09A_FFM01b	485	141.31	521.96	663.27
KFM09A_FFM05	1464	242.76	278.89	521.64
KFM09A_ZFMENE0159A	437	62.56	216.03	278.58
KFM09A_ZFMENE1208	1151	115.97	7.92	123.89
KFM09A_ZFMNW1200	1142	130.94	663.54	794.48
KFM09B_FFM01a	427	71.39	131.99	203.38
KFM09B_FFM01b	419	103.67	203.59	307.26
KFM09B_FFM01c	290	54.35	308.01	362.36
KFM09B_FFM01d	127	104.16	411.79	515.96
KFM09B_FFM01e	41	13.50	547.49	560.99
KFM09B_FFM01f	5	18.69	574.12	592.81
KFM09B_FFM02	265	27.18	76.85	104.03
KFM09B_FZMENE1208	847	67.43	9.27	76.70
KFM09B_ZFMENE0159A	275	25.95	105.35	131.30
KFM09B_ZFMENE2320	381	48.92	362.45	411.37
KFM09B_ZFMENE2325A	286	30.46	516.90	547.36
KFM09B_ZFMENE2325B	119	12.62	561.09	573.71
KFM10A_FFM03a	642	129.63	144.91	274.54
KFM10A_FFM03b	528	145.52	283.25	428.77
KFM10A_ZFMA2	484	64.57	429.15	493.72
KFM10A_ZFMENE2403	77	7.99	274.82	282.80
KFM10A_ZFMWNW0123	981	81.88	62.90	144.79

Classification into SFMs

	The O-3 grid in Figure 5-1b $\Delta_{max}=1.66$ 7 classes	The O-3 grid in Figure 5-1b $\Delta_{max}=1.35$ 9 classes	The alternate O-3 grid in Figure 6-17 $\Delta_{max}>1.66$ 7 classes
	Figure 6-14	Figure 6-15	Figure 6-19
KFM01C_FFM01b	1	1	D
KFM02B_DZ5	1	1	D
KFM01C_ZFMENE0060A	1	6-1	D
KFM07B-Dza	1	1	E
KFM09A_ZFMNW1200	1	6-1	E
KFM04A_ZFMA2	1	1	D
KFM01C_FFM01a	1	1	D
KFM01C_ZFMENE1192	1	1	D
KFM01C_ZFMENE0060C	1	1	D
KFM10A_ZFMA2	2	6-1	E
KFM09B_FZMENE1208	2	2	E
KFM01C_8ZFMA2	2	2	E
KFM06C_DZ1	2	6-1	C
KFM07B-Dzc	2	2	E
KFM01C_FFM02a	2	2	E
KFM01C_FFM02b	2	2	E
KFM01C_FFM02c	2	2	E
KFM02B_DZ3	2	2	E
KFM07A-Dza	2	6-1	E
KFM01B_ZFMA2	2	2	E
KFM01B_FFM02a	2	6-1	E
KFM01D_PDZa	2	6-1	C
KFM09B_FFM02	2	6-1	E
KFM09A_ZFMENE1208	2	2	E
KFM07B-FFM02a	2	6-1	E
KFM07B-Dzb	2	2	C
KFM07B-FFM01b	3	3	F
KFM02B_RU1f	3	3	F
KFM06A_FFM06b	3	3	F
KFM02B_RU1g	3	3	F
KFM02A_DZ3	4	4	C
KFM02A_DZ6	4	4	C
KFM06B_FFM02b	4	4	C
KFM06B_FFM02a	4	4	C
KFM08D-Bsa	4	4	C
KFM06A_DZ1	4	4	C
KFM05A_FFM02	4	7-1	C
KFM09A_FFM01a	4	4	C
KFM08B-FFM02	4	4	C
KFM02A_DZ2	4	4	C
KFM09B_FFM01a	4	7-1	C
KFM01D_FFM02	4	4	C
KFM09B_FFM01b	4	4	C
KFM01A_FFM02a	4	4	C
KFM07B-FFM02b	4	4	C
KFM10A_FFM03b	4	4	C
KFM06C_FFM01	4	4	C
KFM07A-FFM01b	5	5	G

	The O-3 grid in Figure 5-1b Δ_{max}=1.66 7 classes	The O-3 grid in Figure 5-1b Δ_{max}=1.35 9 classes	The alternate O-3 grid in Figure 6-17 Δ_{max}>1.66 7 classes
	Figure 6-14	Figure 6-15	Figure 6-19
KFM01A_FFM01c	5	5	B
KFM02A_FFM01b	5	5	G
KFM02B_RU1d	5	5	G
KFM02A_FFM01a	5	5	G
KFM09B_FFM01f	5	5	B
KFM07B-FFM01a	5	5	G
KFM08A-DZb	6	6-2	A
KFM04A_ZFMNE1188	6	6-2	A
KFM04A_ZFMNW1200	6	6-1	E
KFM07C-Dzb	6	6-1	C
KFM02B_RU1b	6	6-1	C
KFM02B_DZ6	6	6-1	E
KFM08C-DZ2	6	6-1	D
KFM04A_FFM04	6	6-1	D
KFM07B-Dzd	6	6-1	C
KFM08B-DZ1	6	6-1	E
KFM08C-DZ1	6	6-2	C
KFM09A_FFM05	6	6-1	C
KFM09A_ZFMENE0159A	6	6-2	C
KFM02B_DZ2	6	6-1	C
KFM01C_FFM02d	6	6-1	E
KFM07A-DZd	6	6-2	A
KFM07A-DZc	6	7-1	A
KFM01B_ZFMNNW0404	6	6-2	A
KFM09B_ZFMENE2325A	6	6-2	A
KFM06B_DZ1	6	6-1	C
KFM10A_FFM03a	6	6-2	C
KFM01A_ZFMENE1192	6	7-1	C
KFM10A_ZFMWNW0123	6	6-1	E
KFM02B_DZ1	6	6-1	C
KFM06C_DZ5	6	6-1	C
KFM06C_DZ3	6	6-2	C
KFM08D-BSc	6	6-1	C
KFM08D-BSf	6	6-1	C
KFM06C_FFM06b	6	7-1	C
KFM01D_ZFMENE0061	6	6-2	C
KFM10A_ZFMENE2403	6	6-1	D
KFM08A-DZ1	6	6-1	C
KFM07C-FFM01c	6	7-1	C
KFM07C-FFM01b	6	6-2	C
KFM06A_DZ7	6	6-2	A
KFM06A_DZ2	6	6-1	C
KFM09B_ZFMENE0159A	6	6-1	C
KFM09B_ZFMENE2320	6	6-2	A
KFM01D_PDZb	6	6-2	A
KFM08C-DZ3	6	6-1	A
KFM08D-BSb	6	6-1	C
KFM05A_DZ3	6	6-2	A
KFM05A_DZ4	7	7-1	G
KFM07B-FFM02d	7	7-1	C
KFM05A_FFM01a	7	7-2	G
KFM07C-FFM01a	7	7-2	C

	The O-3 grid in Figure 5-1b Δ_{max}=1.66 7 classes	The O-3 grid in Figure 5-1b Δ_{max}=1.35 9 classes	The alternate O-3 grid in Figure 6-17 Δ_{max}>1.66 7 classes
	Figure 6-14	Figure 6-15	Figure 6-19
KFM07A-FFM01a	7	7-2	G
KFM06A_FFM01a	7	7-2	C
KFM06A_FFM01b	7	7-2	G
KFM04A_FFM01	7	7-1	C
KFM06A_DZ10	7	7-2	A
KFM01A_FFM01b	7	7-2	B
KFM08B-FFM01	7	7-1	C
KFM02B_RU1h	7	7-2	G
KFM02B_RU1e	7	7-1	G
KFM02B_RU1c	7	7-1	G
KFM05A_DZ2	7	7-1	A
KFM08C-FFM01a	7	7-2	G
KFM06C_FFM06a	7	7-1	C
KFM08C-FFM01c	7	7-1	G
KFM06A_DZ11	7	7-1	C
KFM07B-FFM02c	7	7-1	A
KFM02B_RU1a	7	7-2	G
KFM08C-FFM06	7	7-1	G
KFM08C-FFM01d	7	7-1	G
KFM08D-Bse	7	7-1	A
KFM08C-FFM01b	7	7-1	C
KFM02A_FFM03b	7	7-2	G
KFM09A_FFM01b	7	7-1	C
KFM02A_FFM03a	7	7-1	6
KFM02A_DZ8	7	7-1	A
KFM02A_DZ7	7	7-1	6
KFM02B_RU2a	7	7-1	6
KFM08A-FFM01	7	7-1	A
KFM01D_FFM01d	7	7-2	6
KFM09B_FFM01c	7	7-1	C
KFM09B_FFM01d	7	7-2	6
KFM09B_FFM01e	7	7-2	6
KFM01D_FFM01c	7	7-2	6
KFM01D_FFM01b	7	7-2	6
KFM01D_FFM01a	7	7-2	6
KFM05A_FFM01b	7	7-2	6
KFM05A_FFM01c	7	7-2	6
KFM01B_FFM02b	7	7-2	C
KFM09B_ZFMENE2325B	7	7-1	A
KFM01B_FFM01b	7	7-2	6
KFM01B_FFM01a	7	7-2	6
KFM01A_ZFMENE2254	7	7-1	6
KFM07A-DZb	7	7-1	C
KFM01A_FFM02b	7	7-1	C
KFM01A_FFM01a	7	7-2	6
KFM06C_FFM05	7	7-1	A
KFM06A_FFM06a	7	7-1	6