# Practicalities of extrapolating 1D fault and fracture size-frequency distributions to higher dimensional samples

Hilde G. Borgos
Department of Mathematical Sciences, The Norwegian University of Science and Technology, 7491 Trondheim, Norway.

Patience A. Cowie, Nancye H. Dawers
Department of Geology and Geophysics, The University of Edinburgh, Grant Institute, West Mains Road, Edinburgh EH9 3JW, U.K.

**Abstract**

Previously published theory, which extrapolates fault and fracture population statistics observed in a one dimensional sample to two and three dimensional populations, is found to be of limited value in practical applications. We demonstrate here how significant the discrepancies may be and how they arise. There are two main sources for the discrepancies: (1) deviations from ideal spatial uniformity (spatial Poisson process) of a fault or fracture pattern, and (2) non-power law scaling of the size frequency distributions of the population. We show that even small fluctuations in spatial density, combined with variance in the estimator of population statistics, can lead to considerable deviations from the theoretical predictions. Ambiguity about power law scaling or otherwise of the underlying population is a typical characteristic of natural data sets and we demonstrate how this can affect the extrapolation of one-dimensional data to higher dimensions. In addition, we present new theoretical approaches to the problem of extrapolation when clustering of faults and fractures is explicitly considered. Clustering is commonly observed in the field as en echelon arrays of fault or fracture segments and we show how this property of natural patterns can be quantified and included in the theory. These results are relevant to building more realistic three dimensional models of the physical properties of fractured rocks, such as fracture permeability and seismic anisotropy.

# 1 Introduction

It has previously been recognised that the observed size-frequency distribution of natural fault patterns depends on the sampling method used (e.g., Marrett & Allmendinger, 1991). For example, if the underlying fault population forms a three dimensional network, but only a two dimensional sample (e.g., a map) is analysed, then many faults will be unrepresented in the resulting size-frequency distribution. Similarly, if the sample comes from a one dimensional traverse (e.g., a borehole) then under-sampling will be even more pronounced. It is only when the dimension of the sample equals the dimension of the underlying fault pattern that a true estimate of the distribution can be determined.

A standard way of making a correction for under-sampled distributions is to assume that faults are independently, uniformly distributed in space with a constant spatial fault intensity (Heffer & Bevan, 1990; Marrett & Allmendinger, 1991; Piggott, 1997; Berkowitz & Adler, 1998). This means that the probability of sampling a fault depends only on its size (e.g., length) and not on its spatial position. Moreover it implies that the size of a fault and its position relative to other faults within the overall pattern are unrelated. There are many field observations that indicate that these assumptions are unrealistic (e.g., Bour & Davy, 1999), as we review in more detail below.

The reason that the assumption of zero spatial correlation has proved so persistent in the literature is that it leads to a very simple result: for a self-similar fault population following a power law size-frequency distribution, as is commonly observed, the difference between the true power law exponent and that of a sample is an integer (Marrett & Allmendinger, 1991). The value of the integer is given by the dimension of the underlying fault network minus the dimension of the sample that is considered. Thus the integer takes the value zero if the pattern is fully sampled; the maximum integer value is 2.0, which corresponds to a 1D sample (i.e., a traverse) of a 3D pattern. In practical applications this simple result is potentially very useful: for example, borehole data on fault or fracture density may be extrapolated to construct a 3D structural network within a rock volume for which bulk properties such as permeability, brittle strain, or seismic velocity and anisotropy are required (Marrett, 1996; Berkowitz & Adler, 1998).

The aim of this paper is to show that, although an integer difference between a sample and the underlying population distribution can be theoretically derived for a reasonable model (Marrett & Allmendinger, 1991), in practice the difference appears to be non-integer. We find that significant deviations from the ideal case result from two main sources: a non-uniform spatial pattern (i.e., clustering) and/or a non-power law size-frequency distribution. We consider both effects here. Our results have important implications for extrapolating natural data sets because of the power law scaling properties. This is because even moderate changes in the value of the power law exponent can result in order of magnitude variations in the numbers of fault or fractures which are predicted. Berkowitz & Adler (1998) found limited success in reproducing natural fracture patterns when assumptions concerning spatial uniformity and power law size scaling were strictly adhered to. Furthermore, Hatton *et al.* (1993) found that the theoretical results of Marrett & Allmendinger (1991) could not explain their laboratory measurements on the relationship between the 3D distribution of microcracks (inferred from acoustic emissions) and the number of microcracks observed in 2D slices of the deformed samples. In spite of these documented failures of the theory, to our knowledge Malinverno (1997) presents the only published attempt to improve or modify the assumptions on which the theory is based. In this paper we address this issue, building on the work of Malinverno (1997), but focusing specifically on sampling of fault populations using information on fault scaling properties from the literature.

In Section 2 of the paper we illustrate that, for a fault pattern that is spatially uniform in 2D, a single 1D line sample (traverse) is sensitive to local variations in fault density and fault sizes. These are natural variations that can lead to deviations from the simple result proposed by Marrett & Allmendinger (1991), when comparing estimated power law exponents. In Section 3 we show how deviations from (ideal) statistical uniformity in the direction parallel to fault strike can be explicitly taken into consideration and quantified. Section 4 then addresses the problem of assuming that the underlying fault size frequency distribution is power law, when sampling problems can result in a distribution that is indistinguishable from an exponential. In the final section (Section 5), spatial clustering perpendicular to fault strike is taken into consideration to quantify the influence of en echelon arrangements of fault segments. Although field observations have indicated that the criteria used for linking fault segments can have a significant effect on fault population analysis (e.g., Wojtal, 1996), this has not previously been quantified theoretically.

## 2 Uniform Spatial Distribution of Faults

A number of field studies of fault populations conclude that the size-frequency distribution of faults is best represented by a power law distribution (e.g., Special Issue of Journal of Structural Geology, Vol. 18, No. 2/3, 1996, and references therein). Let $X$ represent the size of a fault, measured for example as the length $L$ or the maximum displacement $D_{\max}$. The probability density function (pdf) of a power law (or fractal) distribution is given as the pdf of a Pareto distribution (Johnson *et al.*, 1994)

$$f(x) = \frac{\beta x_0^\beta}{x^{\beta+1}}, \qquad x \geq x_0,\ x_0 > 0 \tag{1}$$

where $x_0$ is the minimum fault size. The size-frequency distribution of a fault population is characterised by the power law exponent $\beta$, and this exponent can be estimated based on observed fault patterns. Faults are geometric objects embedded in three dimensions. However, observations are usually obtained from a lower dimension. For fault sizes observed in different dimensions, the value of the power law exponent $\beta$ differs. Let $C_1$, $C_2$ and $C_3$ denote the values of the power law exponent $\beta$ of the length distributions obtained in one, two and three dimensions respectively. Observations of faults in one dimension generally consist of single displacement measurements with no measurable length. In this case the length distribution is obtained from the displacement distribution of the faults being intersected by the traverse, as displacement and length are related (Watterson, 1986).

Due to the under-representation of smaller faults in lower dimensions, the exponents differ according to $C_1 < C_2 < C_3$.

Fault patterns can be represented as marked point processes (Stoyan *et al.*, 1995) where the fault positions form the point process and the marks are given as the fault sizes. In a Poisson point process with independent marks the fault positions are independently, uniformly distributed in space and fault size is independent of position. Under this assumption the following relationships between power law exponents in one, two and three dimensions are found:

$$\begin{aligned} C_2 &= C_3 - 1 \\ C_1 &= C_3 - 2 = C_2 - 1 \end{aligned} \tag{2}$$

(Heffer & Bevan, 1990; Marrett & Allmendinger, 1991; Piggott, 1997; Berkowitz & Adler, 1998). See also Appendix A, where the relationship between 1D and 2D is derived for a generalised size frequency distribution. If $C_1$ or $C_2$ is estimated from observations, the power law exponent in a higher dimension is easily obtained by inversion of Eq. (2) as long as the assumption of spatial uniformity holds. If a relationship $D_{\max} \propto L^n$ is assumed between length and maximum displacement of faults, the power law exponents of fault displacements are given as $C_i' = C_i/n$. The relationships between the power law exponents $C_1'$, $C_2'$ and $C_3'$ for displacements in one, two and three dimensions are then given as

$$\begin{aligned} C_2' &= C_3' - \frac{1}{n} \\ C_1' &= C_3' - \frac{2}{n} = C_2' - \frac{1}{n}. \end{aligned} \tag{3}$$

In uniform rock types and tectonic settings the relationship between $L$ and $D_{\max}$ is found to be linear; $n = 1$ (Cowie & Scholz, 1992a, b; Dawers *et al.*, 1993). Schlische *et al.* (1996) find that the "global" dataset is also best fit by $n = 1$. In this case the power law exponents of the displacement distributions equal the exponents of the length distributions, and Eq. (3) coincides with Eq. (2). For a general $n$, we use the relationship $D_{\max} \propto L^n$ to derive length distributions from displacements measured along 1D traverses.

Using Eq. (2) or (3) to extrapolate from one to two dimensions, or from two to three, can give unreliable results of the power law exponent in higher dimensions. Obviously, if the assumptions leading to the results are not fulfilled then the theoretical relationships will change. However, when power law exponents are estimated, the uncertainty lies not only in the fulfilment of the requirements behind Eq. (2) and (3), but also in the validity of the estimate. The variance of the estimator is a source of uncertainty, regardless of the spatial distribution of fault positions. This is illustrated through a simulation example. A 2D fault pattern consisting of 25,000 faults is simulated. A small section of the fault pattern is given in Figure 1a. The fault pattern is generated using a uniform distribution of fault centre points, a uniform distribution of orientations on the interval $[0°, 360°]$, and for the lengths a power law distribution with exponent $C_2 = 1.6$ is used. To study the length distribution obtained in 1D, data is collected along 1000 horizontal traverses at randomly chosen positions. For each traverse the power law exponent of the length distribution is estimated from the intersected fault traces, using a maximum likelihood estimator. The maximum likelihood estimator is known to be biased (Clark *et al.*, 1999). Figure 1b shows a histogram of the estimated exponents $C_1$.

The assumptions behind the result in Eq. (2) are fulfilled, and the theoretically predicted value of the power law exponent for lengths observed along the traverses is $C_1 = 0.6$. The estimated exponents are centred around the theoretical value, although slightly biased with an average of 0.62. However, the important point shown in Fig. 1b is that the standard deviation is 0.07 and that the bulk of the estimates lies in the range 0.5–0.75, giving a difference between $C_2$ (=1.6) and $C_1$ of 0.85–1.1. Thus in these examples the natural variations in the estimator can give a deviation of 15–25% between the estimated exponent and the theoretically predicted value. Similar conclusions are reached if other fault orientations are considered, from uniformly distributed orientations over the interval $[0°, 360°]$ to all faults being parallel. The key issue here is the number of faults intersected by the traverse. In the simulation example, the number of faults intersected by the traverse lines varies considerably from 60 to 105 with a mean of 82. For traverse datasets containing a larger number of faults the bias and standard deviation will be lower, while for a decreased number of observations the uncertainty increases.

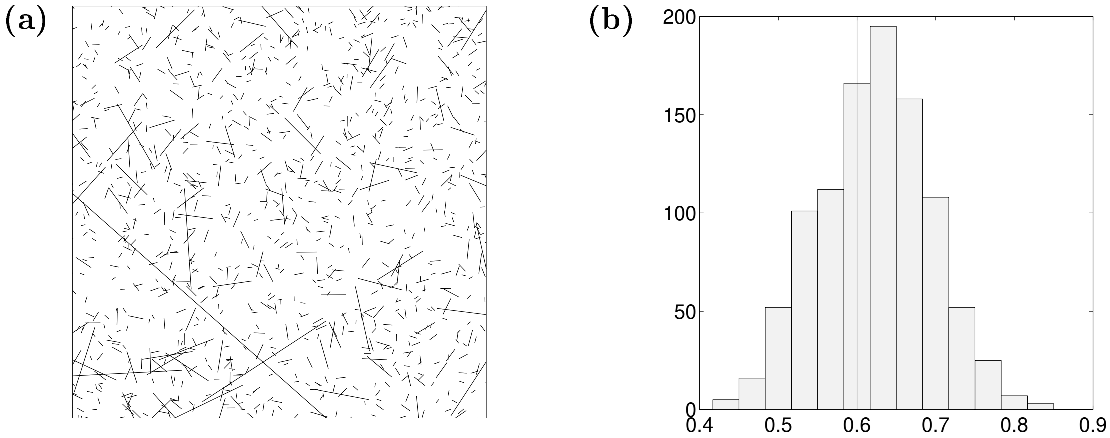


Figure 1: (a) Simulated fault pattern with a uniform spatial distribution and a power law length distribution with exponent $C_2 = 1.6$. (b) Histogram of estimated exponents $C_1$ from 1000 traverses. The solid line marks the theoretical value $C_1 = C_2 - 1 = 0.6$.

Clark *et al.* (1999) compare estimators for the power law exponents, and discuss inefficiencies in the geoscientists' traditional estimator. The standard way of estimating the power law exponent is by fitting a straight line to the plot of $\log N(x)$ versus $\log x$, where $N(x)$ is the number of faults larger than $x$. The estimates presented in Fig. 1b are obtained using a maximum likelihood estimator, which has known mean and variance (Clark *et al.*, 1999). For $C_1 = 0.6$ and a number of observations ranging from 60 to 100, the standard deviation of the maximum likelihood estimator ranges from 0.08 to 0.06, in agreement with the empirical standard deviation of 0.07, found above. If the number of observations was increased to $\sim 1000$, the standard deviation of the maximum likelihood estimator would decrease to 0.02, reducing the error in the estimated power law exponent to $\leq 5\%$. But such large samples are unusual in natural data sets; samples sizes of tens to hundreds are much more typical.

# 3 Deviations from a Uniform Spatial Distribution

Two main assumptions behind the results in Eq. (2) and (3) are uniform spatial distribution of fault positions and power law size-frequency distribution. In this section deviations from a uniform spatial distribution are considered. The relationship $C_1' = C_2' - 1/n$ observed in a traverse through a two dimensional fault pattern should be independent of the position of the traverse if the assumption of independently, uniformly distributed fault positions holds. Deviations from the uniform spatial distribution will alter the relationship between $C_1'$ and $C_2'$.

We use a similar approach to Malinverno (1997), who studied the statistical distribution of turbidite beds, and the relationship between the bed thickness distribution observed along a 1D traverse (a borehole) and the overall bed extent in 3D. Malinverno (1997) idealised turbidite beds as conical 3D bodies with a maximum thickness in the centre, decreasing to zero thickness laterally. In this work we assume that turbidite bed thickness is analogous to variations in displacement over a fault plane, where the displacement typically is decreasing from a maximum near the centre to zero at the fault boundaries. Malinverno (1997) considers vertical boreholes through horizontal sedimentary layers, while the fault data studied in this work are 1D transects across a map (Fig. 2). The geometric representation of turbidite bed thickness used by Malinverno (1997) is used here to represent the displacement profile along a fault trace in 2D. When the traverse intersects a fault, it intersects at an arbitrary distance $r$ from the centre point. The displacement observed in the traverse is then given by $d(r; D_{\max})$, where $D_{\max}$ denotes the maximum displacement. We can write the relationship between length and maximum displacement as $D_{\max} = \gamma L^n$. The profile is then given as

$$d(r; D_{\max}) = D_{\max} \left\{ 1 - \left( 2r \left( \frac{\gamma}{D_{\max}} \right)^{1/n} \right)^{\alpha} \right\}, \qquad 0 \leq r \leq \frac{1}{2} \left( \frac{D_{\max}}{\gamma} \right)^{1/n} \tag{4}$$

where the parameter $\alpha > 0$ controls the shape of the displacement profile. For $\alpha < 1$ the displacement profile is peaked at the centre, for $\alpha = 1$ the profile is triangular while for $\alpha > 1$ the displacement increases rapidly near the end points and flattens out in the centre part; see also (Malinverno, 1997). As $\alpha \to \infty$ the profile becomes rectangular. The displacement is assumed to be at its maximum at the centre point of the fault, and all faults are assumed to be approximately parallel with an orientation perpendicular to the traverse.

We now consider a non-uniform spatial distribution of faults, where the location of the traverse has influence on the observed power law exponent $C_1'$. Malinverno (1997)

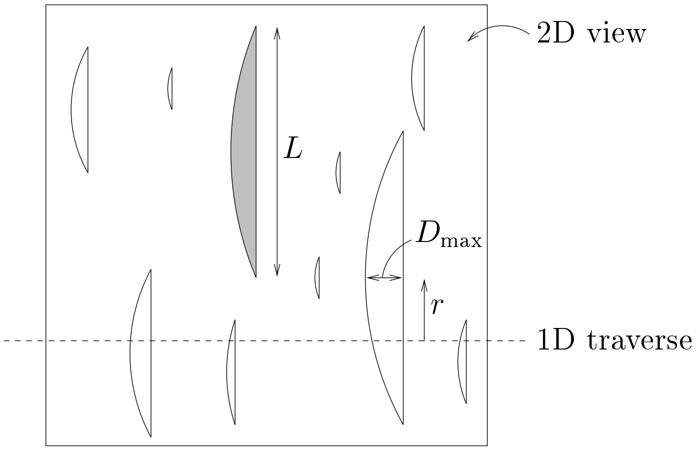


Figure 2: Illustration of 1D traverse line through a 2D fault pattern. The shaded area indicates the fault displacement profile of a fault of length $L$, Eq. (4). The displacement is at a maximum $D_{\max}$ at the fault centre point, and the distance from the traverse line to the centre point of an intersected fault is $r$.

introduces a distribution $f(r) \propto r^{\delta-1}$, $\delta > 0$, for the distance from a sampling line to the centre point of a turbidite bed. A similar distribution can be used for the distance from the traverse to a fault centre point, measured perpendicular to the traverse. If $\delta = 1$ the fault positions are independently, uniformly distributed. If $\delta < 1$ the spatial distribution relative to the traverse deviates from the uniform distribution, and the faults are clustered in a self-similar fashion in the direction of the strike. The distribution $f(r)$ represents a decreased intensity of fault positions away from the traverse when $\delta < 1$, which can be interpreted as the traverse being located in a high intensity area of faults compared to the overall fault density. Reformulating the results of Malinverno (1997), the power law exponent of displacements observed along the traverse is found to be

$$C_1' = C_2' - \frac{\delta}{n}. \tag{5}$$

The corresponding result for power law exponents $C_i = nC_i'$ of fault lengths is

$$C_1 = C_2 - \delta. \tag{6}$$

This result shows that if the spatial distribution perpendicular to the traverse line is not uniform, but scales in a self-similar fashion, the difference between power law exponents in two and one dimension differs from the theoretically obtained value in Eq. (2) and (3). For a non-uniform spatial distribution with $\delta < 1$, the difference between $C_2$ and $C_1$ will be less than the expected value 1. For $\delta = 1$ Eq. (6) and (5) reduces to Eq. (2) and (3) as expected. The relationships in Eq. (6) and (5) are independent of the shape parameter $\alpha$, thus the choice of $\alpha$ is not critical, although if $n = 1$ then the fault population is self-similar with a constant value of $\alpha$. For interacting and linked faults $\alpha$ is no longer a constant (Willemse *et al.*, 1996), but the effect of this variation may be of secondary importance.

# 4 Deviations From Power Law Size-Frequency Distribution

We now examine the second major assumption that underlies Eq. (2) and (3), that of the power law size-frequency distribution. Many field data sets, in continental regions at least, have been interpreted as showing power law size scaling (see Special Issue of Journal of Structural Geology, Vol. 18, No. 2/3, 1996). However, any natural data set is affected to a greater or lesser degree by the effects of limited resolution and censoring. The former leads to under-sampling of small faults while the latter leads to inadequate sampling of faults that are comparable in size to the dimensions of the sampling domain (e.g., a map or cross-section). Although various corrections have been proposed to minimise and quantify these effects (e.g., Pickering *et al.*, 1995, 1996), the corrections assume that the true distribution is power law. However, some fault populations, such as those formed at mid-ocean ridges (Cowie *et al.*, 1994) are better described by an exponential distribution and Ackermann *et al.* (1997) and Spyropoulos *et al.* (1999) have documented both power law and exponential distributions in analogue experiments of fault growth (see also Ackermann, Withjack & Schlische, The effects of mechanical layer thickness on the systematics of an evolving population of normal faults in experimental models, manuscript submitted to J. Struct. Geol., September 1999).

A general case of deviations from a power law distribution of fault size will not be considered, but the special case of fitting a power law exponent to data that is in fact exponentially distributed is discussed, as this is a common ambiguity in interpreting fault population data (Ackermann *et al.*, 1997; Cowie, 1998). The relationship between the length distributions in two dimensions and along a 1D traverse is studied, to quantify the under-representation of small faults for exponentially distributed data. Note that if fault displacements were to be analysed instead, the displacement profile must also be included in the calculations, giving analytically non-tractable expressions.

Let $N(x)$ be the cumulative number of observed faults with length larger than $x$. If the fault length follows a power law distribution, Eq. (1), a plot of $\log N(x)$ versus $\log x$ gives a straight line with slope $-\beta$. If the fault length follows an exponential distribution with intensity $\lambda =$ (mean of $x)^{-1} = <x>^{-1}$ and probability density function

$$f(x) = \lambda \exp(-\lambda x), \qquad x \geq 0 \tag{7}$$

the log-log plot produces a concave shape to the curve. However, a similar type of concave curve can be the result of plotting poorly sampled data from a power law distribution, where truncation, censoring and finite range effects have led to a flattening of the curve for small $x$ and a steep fall off for large $x$ (Pickering *et al.*, 1995). Thus data from a truly exponential distribution can mistakenly be assumed to follow a power law, using the centre part of the log-log plot to fit a power law exponent to the data.

Suppose the fault lengths in two dimensions follow an exponential distribution, but a power law exponent $C_2$ is fitted to the data. This exponent is then compared to a power law exponent $C_1$ fitted to data along a one dimensional traverse through the observation area. Since a plot of $\log N(x)$ versus $\log x$ for exponentially distributed data does not give a straight line, the slope of a local line-fit will vary depending on $x$. When a power law exponent is calculated, a straight line is fitted to the centre part of the data, assuming this part to be approximately straight. The fitted power law exponent will be close to the absolute value of the slope of the true curve for this part of the data. At length $x$ the theoretical absolute value of the slope is

$$C_2(x) = \lambda x. \tag{8}$$

If the fault length in two dimensions follows an exponential distribution, Eq. (7), the length of faults intersected by a one dimensional traverse follows a gamma distribution (Johnson *et al.*, 1994)

$$f(x) = \lambda^2 x \exp(-\lambda x), \qquad x \geq 0, \tag{9}$$

see Appendix A. The absolute value of the slope of the curve $\log N(x)$ versus $\log x$ at length $x$ is

$$C_1(x) = \frac{(\lambda x)^2}{\lambda x + 1}, \tag{10}$$

as shown Appendix B. The resulting difference in exponents for the 2D and 1D population at length $x$ is then

$$C_2(x) - C_1(x) = \frac{\lambda x}{\lambda x + 1} < 1, \tag{11}$$

independent of the value of $\lambda$. One of the assumptions leading to the result in Eq. (2) is no longer valid and consequently $C_2 - C_1$ differs from 1. Typical values of fault lengths in the exponential distribution, Eq. (7), in 2D and the corresponding distribution, Eq. (9), in 1D are the mean values $<x>= 1/\lambda$ in 2D and $<x>= 2/\lambda$ in 1D respectively. These choices of $x$ give $C_2(x) - C_1(x)$ values of 1/2 and 2/3 respectively. Thus for typical fault lengths of the given distributions, the difference in slopes between one and two dimensions is significantly smaller than 1.

Based on the result in Eq. (11) it can also be concluded that if the distribution of fault length is neither power law nor exponential, but a mixture of the two distributions (Ackermann *et al.*, 1997; Spyropoulos *et al.*, 1999), the difference in fitted exponents in two and one dimension will be less than 1. Generally these calculations show that if the power law exponent is fitted to data not originating from a power law distribution, the relationships in Eq. (2) and (3) between exponents obtained in different dimensions are no longer valid. Specifically we find that, for the case shown here, $C_2 - C_1 < 1$ and may be as low as 0.5–0.67.

# 5 Interacting and Linked Fault Structures

In two dimensions faults are observed either as relatively isolated structures, or as interacting faults forming segments of en echelon arrays. En echelon arrays are easily detected in

2D observations, and faults may be classified as isolated, interacting or linked using such criterion as the ratio of overlap to separation of the fault tips (Gupta & Scholz, A model of normal fault interaction based on observations and theory, manuscript submitted to J. Struct. Geol., 1999). When faults are observed along a 1D traverse however, it is more difficult to determine whether two structures occurring close in space are two isolated faults or are segments of the same fault structure along strike. Wojtal (1996) showed how criteria used for defining linked fault segments in a data set of thrust faults from the Cumberland Plateau in Tennessee, has an important influence on the estimated power law exponent in an analysis of fault size-frequency distributions.

Figure 3 illustrates the ambiguity in the interpretation of fault linkage on a 1D traverse. Fig. 3a shows the throw profile along a fault array modified from Dawers & Anders (1995), and Fig. 3b gives a map view of the fault array showing the location of two traverses $x - x'$ and $y - y'$ where the total throw is comparable. Fig. 3c compares the two 1D traverses across the fault. Along $x - x'$ four separate structures are seen across a zone $\sim 500m$ wide, whereas along $y - y'$ a single fault plane is observed. Without 2D information it is not possible to judge from traverse $x - x'$ alone whether the four separate faults are linked along strike, even though it is clear from the summed displacement profile that they do form part of a single larger structure.

One of the main assumptions behind the results in Eq. (2) and (3) is not satisfied for fault segments in en echelon arrays, since these represent a type of clustering of fault segments. In this section we consider a fault pattern that consists of a mixture of fault arrays and isolated faults. For simplicity it is assumed that the results in Eq. (2) and (3) hold for isolated faults or linked fault arrays separately, and also for the individual segments of the linked systems, but not for the overall pattern.

An en echelon array is composed of a number of shorter fault segments. We separately define the lengths of the individual segments and the length of each fault array. If the length distribution of arrays is compared to the length distribution of the single fault segments, it is obvious that the length distribution of the single segments has a larger proportion of small faults than that of the linked arrays. Suppose both length distributions are power law distributions, with exponent $\beta_S$ for the single segments and $\beta_A$ for the arrays. The difference in the amount of large faults implies $\beta_S > \beta_A$. The same argument applies for the displacement. The displacement of an linked fault array is a sum of displacements on the smaller fault segments, thus the linked structures tend to have larger displacements than the shorter single segments which the array is composed of.

In this work we distinguish between isolated faults, fault arrays and single fault segments which comprise the linked systems. Strictly speaking the segments making up a linked fault system have displacement profiles differing from the typical profile of an isolated fault, while the shape of the overall (summed) displacement profile of the array is comparable to that of isolated faults (e.g. Peacock & Sanderson, 1991; Dawers & Anders, 1995; Willemse *et al.*, 1996). For simplicity it is assumed in this work that isolated faults and single segments in a linked fault system have the same size-frequency distribution, whether we compare either length or displacement. Consequently we are also making the assumption that the maximum size of the isolated faults are smaller than that of the fault

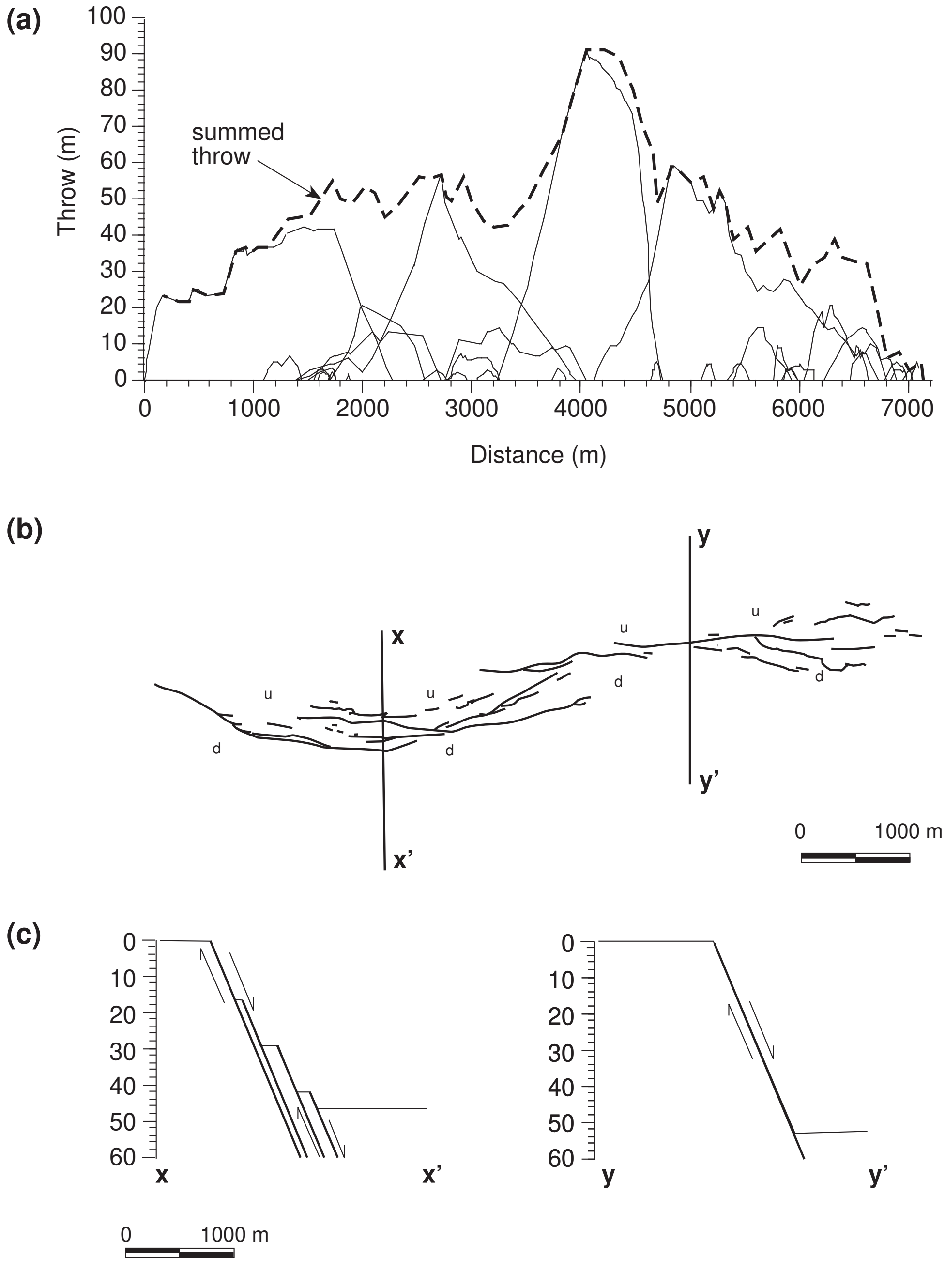


Figure 3: (a) Throw profile along a linking fault array from the Volcanic Tableland in eastern California modified from Dawers & Anders (1995). Solid lines show throw variations along individual segments of the array, dashed line shows the total throw along the entire array obtained by summing together the throws of the separate segments. (b) Map view of fault array showing the location of two traverses $x - x'$ and $y - y'$ shown in (c). (c) Comparison of the two 1D traverses $x - x'$ and $y - y'$ across the fault where the total throw is comparable. See text for further discussion.

arrays, and that the populations of isolated faults and single segments are thus comparable.

Let $X$ represent the maximum displacement of a fault observed in two dimensions, and suppose $X$ follows a power law distribution. The power law exponent depends on $X$ representing an isolated fault or an en echelon array. If $X$ represents an isolated fault the power law exponent is $\beta_S$, and the pdf of $X$ is denoted $f(x;\beta_S)$. If $X$ represents a linked fault structure, the power law exponent is $\beta_A$ and the pdf is $f(x;\beta_A)$. The overall size-frequency distribution of fault displacements observed in two dimensions is a combination of the two distributions:

$$f(x) = (1-\theta)f(x;\beta_S) + \theta f(x;\beta_A) = (1-\theta)\cdot\frac{\beta_S x_0^{\beta_S}}{x^{\beta_S+1}} + \theta\cdot\frac{\beta_A x_0^{\beta_A}}{x^{\beta_A+1}}, \qquad 0\le\theta\le 1 \quad (12)$$

where $\theta$ is the proportion of linked fault structures in the fault pattern, given as the number of linked fault structures divided by the total number of faults. The parameter $\theta$ is determined by imposing a linkage criterion on the 2D fault observations. A simple criterion of linkage is to classify two faults as linked based on the value of the overlap-to-separation ratio $O/S$ (Fig. 4). The range of values of $O/S$ of linked faults typically lies in the interval 1–10 (Aydin & Schultz, 1990; Huggins *et al.*, 1995; Gupta & Scholz, 1999, Schultz, in press, Dawers, Malinverno & Scholz, Fault displacement and length distributions, linkage and strain on the Volcanic Tableland, California, manuscript in preparation).

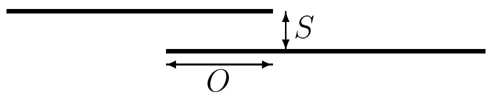


Figure 4: Definition of overlap $O$ and separation $S$.

The value of $\theta$ depends on the linkage criterion, and also on the resolution of the data. Let $N_I$ be the number of isolated faults at a given data resolution, and let $N_A$ be the number of linked fault arrays. The value of $\theta$ is then given as

$$\theta = \frac{N_A}{N_I + N_A}. \quad (13)$$

Figure 5 gives an illustration of three fault patterns with a varying value of $\theta$. For a normal fault population in the Volcanic Tableland of eastern California (Dawers *et al.*, 1993; Dawers & Anders, 1995), the proportion $\theta$ is thought to be less than 0.5 (Dawers *et al.*, in prep.). For this field data, the values of $O/S$ in our linkage criterion lie in the interval 3–10 (Dawers, 1996), as indicated by asymmetric displacement profiles of interacting faults (Willemse *et al.*, 1996).

Figures 6, 7 and 8 give an illustration of the combined distribution function (12) and the appearance of data originating from this distribution. The observations in Figures 6 and 7 are generated using statistical sampling techniques. Figures 6 and 7 show log-log plots of relative cumulative number $N(x)/N_T$ versus size $x$, where $N_T$ is the total number of faults. The expression $N(x)/N_T$ is an estimate of the cumulative distribution

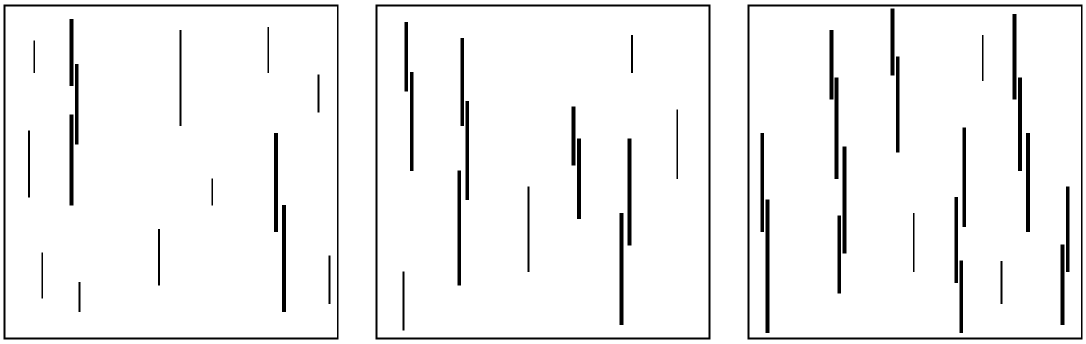


Figure 5: Three fault patterns with different proportion $\theta$ of linked fault structures. The thick solid lines represent linked faults with $O/S$ ratios in the range 3–8.

function $S(x) = \text{Prob}\{X \geq x\}$, see Appendix B. Figure 8 shows the cumulative distribution functions $S(x)$ for different values of $\theta$.

The combined distribution is only a power law distribution in the extremes $\theta = 0$ or $\theta = 1$, where it equals one of the two separate distributions $f(x; \beta_S)$ or $f(x; \beta_A)$. For intermediate values, $0 < \theta < 1$, Eq. (12) does *not* give a new power law distribution. Thus a plot of $\log N(x)$ versus $\log x$ does not give a straight line according to the theory. However, as Figure 6 shows, data from the distribution $f(x)$ appear to be approximated quite well by a straight line on log-log plot. Figure 6 shows a log-log plot of an increasing number of observations generated from the distribution $f(x)$, using power law exponents $\beta_S = 1.8$ and $\beta_A = 1.3$. If these were real observations with the theoretical expression in Eq. (12) unknown, most geologists would conclude that the data originate from a power law distribution (Scholz *et al.*, 1993).

Figure 7 compares the combined dataset in Fig. 6b with two datasets generated from the original power law distributions with exponents $\beta_S$ and $\beta_A$. On a cumulative log-log plot, the combined distribution is located between the two distributions of which it is composed. Figure 8 shows a log-log plot of the theoretical cumulative distribution function $S(x)$ for the combined distribution, together with the corresponding curves for $f(x; \beta_S)$ and $f(x; \beta_A)$, for an increasing value of the proportion $\theta$. The curve of $\log S(x)$ for the combined distribution is slightly convex. If however the data are assumed to follow a power law distribution, it is observed from Figs. 7 and 8 that an apparent power law exponent $\beta_C$ of the combined distribution $f(x)$ lies between $\beta_A$ and $\beta_S$, thus $\beta_A < \beta_C < \beta_S$. The estimated exponents of the observations in Fig. 6 are plotted as straight lines, together with the curved theoretical line. The curve of $\log S(x)$ in Fig 8b, using $\theta = 0.25$, corresponds to the samples shown in Figs. 6 and 7. As a result of the slightly convex form of the curve given by Eq. (12), the straight line fitted to the observations fall below the theoretical line as the fault size increases (Fig. 6). When the sample size increases, the fitted power law tend to under-estimate the number of large faults (Fig. 6d).

Table 1 gives estimated power law exponents $\beta_C$ for an increasing number of observations and a varying proportion $\theta$ of linked fault structures. Each estimate is based on 10,000 samples generated from the combined distribution (12), using $\beta_S = 1.8$ and $\beta_A = 1.3$, and the mean and standard deviation of the estimates are listed in the table. It is observed from the table that $\beta_C$ approaches $\beta_S$ for small ratios $\theta$, and approaches $\beta_A$ as $\theta$ increases, in agreement with the results shown in Figure 8. It is also observed from Table 1 that as the number of observations decreases, the estimated power law exponent $\beta_C$ increases. This is due to the fact that the theoretical line of $\log N(x)$ is steepest for small fault sizes (Fig. 8). This region of the curve dominates the small datasets, which do not span more than one order of magnitude (Fig. 6). Figure 9 summarises the results from Table 1, for observations consisting of 500 faults. The error bars show the uncertainty in the estimate of the power law exponent $\beta_C$. A larger number of observations would produce similar plots, only with shorter error bars, while a smaller number of observations would give a wider range of uncertainty. Note that the exponent $\beta_C$ appears to decrease approximately linearly with $\theta$.

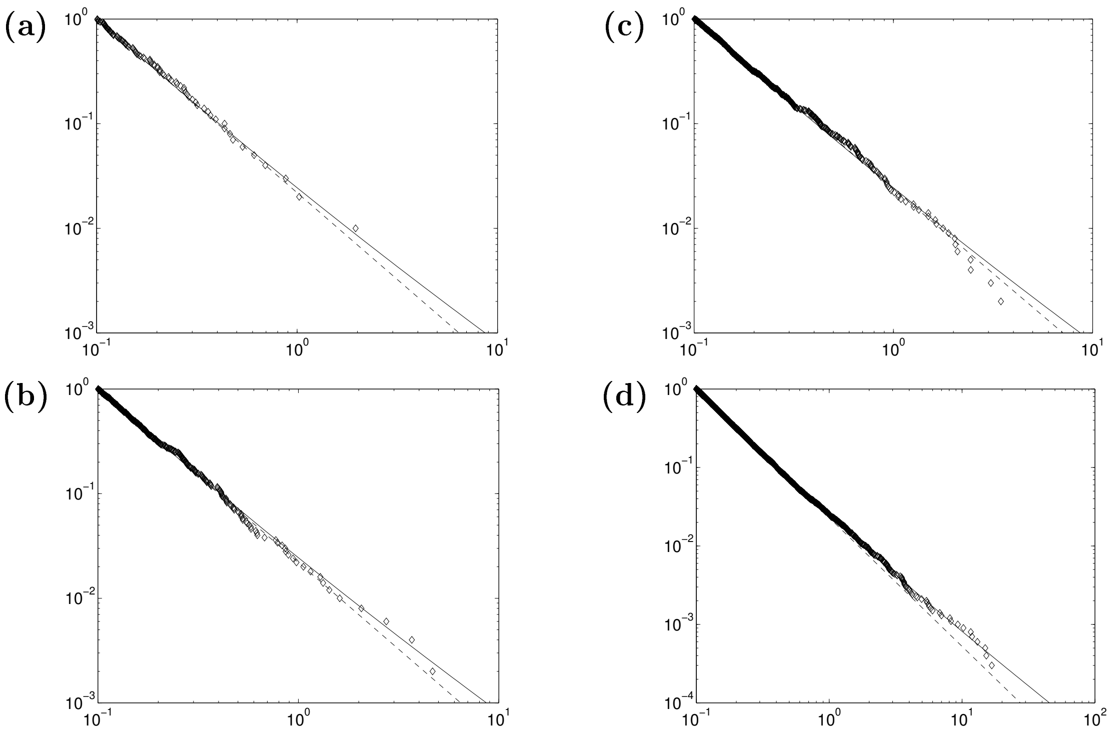


Figure 6: Samples generated from the combined distribution $f(x)$ in Eq. (12), with $\beta_S = 1.8$, $\beta_A = 1.3$ and $\theta = 0.25$. The number of observations in each sample is (a) 100, (b) 500, (c) 1000 and (d) 10,000 and the estimated power law exponents are (a) $\beta_C = 1.66$, (b) $\beta_C = 1.66$, (c) $\beta_C = 1.62$ and (d) $\beta_C = 1.64$. The estimated exponents are plotted as straight dashed lines, and the theoretical lines are plotted as solid lines.

As pointed out above, the distribution $f(x)$ in Eq. (12) is not strictly a power law distribution. Assume however that the deviations from a power law is small enough to give

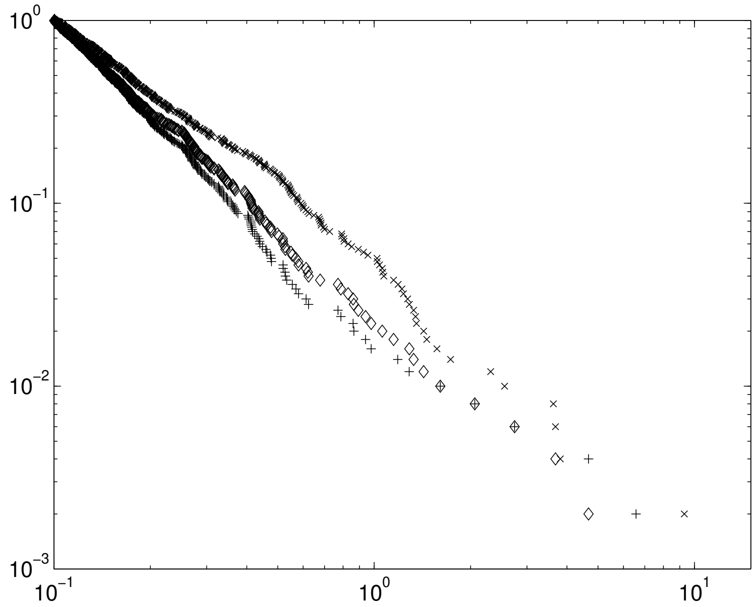


Figure 7: Comparison of the combined dataset in Fig. 6b (plotted as $\diamond$) with two datasets generated from the distributions $f(x;\beta_S)$ (plotted as $+$) and $f(x;\beta_A)$ (plotted as $\times$). The power law exponent of single faults is $\beta_S = 1.8$ and for fault arrays $\beta_A = 1.3$. The estimated power law exponent of the combined distribution is $\beta_C = 1.66$.

| Sample | Ratio $\theta$ of linked fault structures, Eq. (13) | | | | |
|---|---|---|---|---|---|
| size | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 |
| 100 | 1.751 (0.180) | 1.657 (0.173) | 1.525 (0.160) | 1.412 (0.146) | 1.354 (0.139) |
| 500 | 1.736 (0.078) | 1.646 (0.075) | 1.513 (0.070) | 1.399 (0.063) | 1.340 (0.061) |
| 1000 | 1.734 (0.055) | 1.643 (0.054) | 1.511 (0.049) | 1.399 (0.045) | 1.339 (0.043) |
| 10,000 | 1.734 (0.018) | 1.642 (0.017) | 1.510 (0.016) | 1.397 (0.014) | 1.337 (0.014) |

Table 1: Estimated power law exponents $\beta_C$ based on 10,000 generated samples of increasing sample size from the combined distribution in Eq. (12). Different values of the proportion $\theta$ of linked fault structures are used. The table gives the mean value and standard deviation of the estimates, calculated from the 10,000 samples. The power law exponents $\beta_S = 1.8$ and $\beta_A = 1.3$ are used for isolated faults and fault arrays.

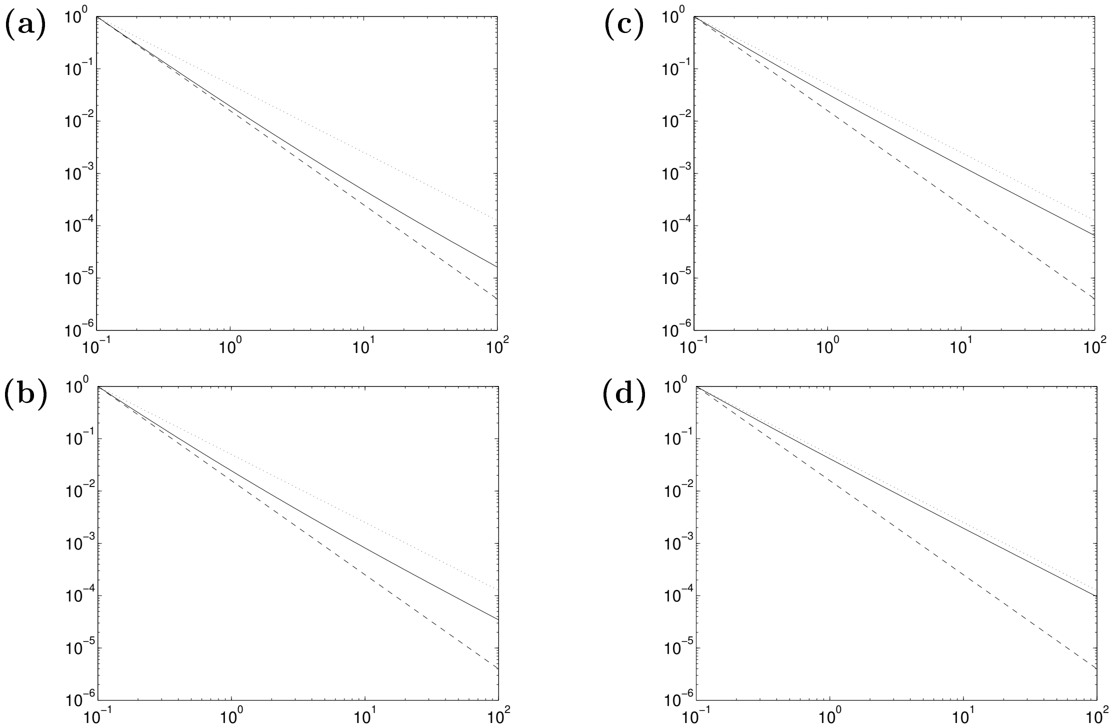


Figure 8: Log-log plot of cumulative distribution function $S(x)$ for the two distributions $f(x;\beta_S)$ (dashed lines) and $f(x;\beta_A)$ (dotted lines) with $\beta_S = 1.8$ and $\beta_A = 1.3$, and for the combined distribution $f(x)$ in Eq. (12) (solid lines). The lines are plotted using the ratios (a) $\theta = 0.1$, (b) $\theta = 0.25$, (c) $\theta = 0.5$ and (d) $\theta = 0.75$.

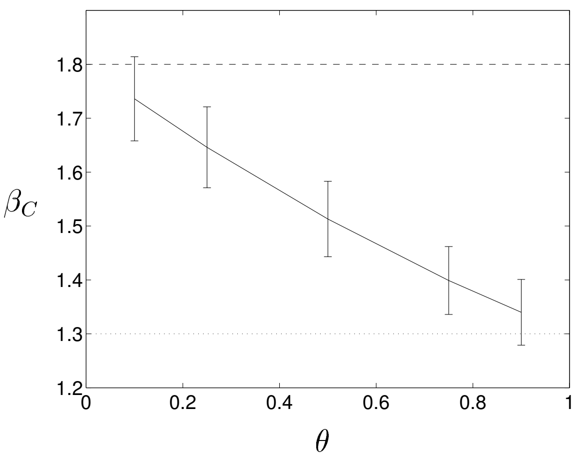


Figure 9: Variations in the estimated power law exponent $\beta_C$ as a function of the ratio $\theta$ of linked fault structures (Tab. 1). The error bars show one standard deviation of the estimator for datasets of size 500 faults. The horizontal dashed line shows the value of the power law exponent $\beta_S = 1.8$, the dotted line the exponent $\beta_A = 1.3$.

$C_1' \approx C_2' - 1/n$ for the apparent power law exponents. The fault pattern in 2D, consisting of a mixture of isolated faults and en echelon arrays, has an apparent power law exponent of $\beta_C$. The corresponding exponent for observations along a traverse should then be $\beta_C - 1/n$, according to Eq. (3). However, if no linkage criterion is used along the traverse, and all observations are treated as single faults, the predicted power law exponent is $\beta_S - 1/n$, since isolated faults and single fault segments in 2D have exponent $\beta_S$. Comparing the 2D data with exponent $C_2' = \beta_C$ and 1D data with exponent $C_1' = \beta_S - 1/n$ thus gives a difference of

$$C_2' - C_1' = \frac{1}{n} - (\beta_S - \beta_C) < \frac{1}{n} \tag{14}$$

since $\beta_S > \beta_C$. In Section 3 clustering along strike was considered, and a new theoretical expression for the difference $C_2' - C_1'$ was obtained. A comparable expression based on Eq. (14) can not be derived, since the exponent $\beta_C$ is only an apparent exponent, not given by an analytical expression. An estimate of $\beta_C$ will depend on the situation: on all parameter values, the ratio $\theta$ and the number of observations (Tab. 1). Thus in order to estimate the difference between $C_2'$ and $C_1'$, an objectively measured 1D sample needs to be re-interpreted (or "pre-processed") using some linkage criterion in an attempt to obtain 1D data corresponding to the distribution in Eq. (12).

There are two possible ways to do this: (1) by taking a purely statistical approach to linkage probability, as outlined in Appendix C, or (2) by re-evaluating the field data from which the 1D data has been derived, as described, for example, by Dawers (1996), Wojtal (1996), Schultz (in press), Dawers *et al.* (in prep.). Concerning the latter approach it is worth noting that for many published examples of 1D fault population data, the size-frequency distribution is obtained for a single stratigraphic horizon within a cross-section (e.g., Walsh *et al.*, 1991, Marrett & Allmendinger, 1992). Thus, there is potentially some two dimensional information that may be extracted by examining the upper and lower tips of the faults which have been measured to assess how many are en echelon with other faults higher or lower within the section. Then using an overlap-separation criterion for linkage in the plane of the cross-section, and assuming that this applies in plan view, those faults belonging to arrays can be separated from those that are isolated to obtain a value for $\theta$. The key assumption here is that evidence for linkage in the cross-sectional view indicates linkage also occurring in plan view; a reasonable starting assumption but one which has not been independently verified.

# 6 Discussion

We have shown that there are several practical situations in which the addition or subtraction of an integer value from the power law scaling exponent does not provide an adequate representation of how populations in 1D, 2D and 3D are related. We have focused on the extrapolation of 1D data to 2D and in particular the effects of spatial variations in fault density and en echelon clustering. These effects lead to a difference between $C_2$ and $C_1$

that is $< 1.0$. The implication of this is that if 1D data are extrapolated to 2D without considering clustering then the relative numbers of small scale faults and fractures are overestimated in the 2D distribution, the relative number of large scale structures is underestimated, and the size of the largest fault is also underestimated. Moreover, not only do you overestimate the numbers of small fractures, the absence of spatial correlation information means that these small fractures are randomly positioned in the extrapolation to a 2D pattern whereas in natural patterns they tend to be aligned in arrays.

Even though 2D data allows some evaluation of en echelon clustering and linkage, the same problems arise in going from 2D to 3D if the morphology of the fault surfaces and the network of faults are three dimensional. Thus the same effects will be seen as we have demonstrated here. Berkowitz & Adler (1998) generated 3D fracture networks from 2D field observations using the established relationships given in Eq. (2). They then compared 2D slices through the model 3D volume with the original field maps, in terms of both the predicted 2D exponent and the visual appearance of the patterns. The original field measurements of fracture length do take into consideration 2D linkage between closely spaced en echelon fractures (Segall & Pollard, 1983). However, in the extrapolation to 3D the spatial locations of the fractures are assumed to be independently, uniformly distributed, although this assumption is not tested (Berkowitz & Adler, 1998). For several cases the calculated exponents agree quite well, and discrepancies can largely be attributed to having small numbers of fractures in the data sets which make the statistics less reliable. This agreement is expected since their 3D simulated fracture patterns fulfil the requirements behind the result $C_3 - C_2 = 1$. However visual comparison of the real and modelled patterns shows two interesting features: Firstly, the real pattern exhibits zones of more concentrated fracturing in the 2D plane separated (perpendicular to strike) by zones that are devoid of fractures, and secondly, strike-parallel alignment and segmentation of the natural fractures is very common. Neither of these properties is captured by the simulations. However, these are exactly the properties which will strongly influence the porosity structure, permeability and seismic velocity of fractured rock volumes.

Marrett (1996) discusses the impact of power law scaling of fault or fracture populations on "aggregate" properties, such as total fracture porosity, permeability, total fracture surface area, total fault strain, and shear-wave anisotropy. He concludes that, except for fracture surface area, larger fractures/faults contribute more to aggregate properties than do the smaller ones for most values of the scaling exponents. However, these conclusions depend on the value assumed for the power law exponents in 3D and on the assumption that the spatial pattern of the fracture or fault population originate from a uniform distribution. The present work has shown that the conclusions of Marrett (1996) may need to be re-evaluated in the light of spatial correlation information which has previously been ignored.

Hatton *et al.* (1993) derived an empirical relationship between 2D and 3D scaling exponents of the form $C_3 = a_1 C_2 + a_2$ where $a_1$ and $a_2$ are constants. According to their empirical analysis $a_1 = 1.3$ and $a_2 = -0.2$, whereas comparison with the present work would suggest $a_1 = 1.0$ and $0 \leq a_2 \leq 1.0$. Hatton *et al.* (1993) obtained this relationship from rock deformation experiments conducted under a variety of dry, damp and wet conditions which were designed to examine the differences in sub-critical crack growth processes in the presence/absence of a fluid. Estimates of $C_3$ were calculated from the size frequency distribution of acoustic emissions produced during the experiments whereas $C_2$ was obtained from populations of cracks observed in 2D slices of the deformed samples. Values of these exponents from both wet and dry experiments were combined together to obtain the relationship between $C_2$ and $C_3$, even though the morphology of the crack populations differed significantly. For the dry experiments the deformation was localised onto longer more continuous cracks, whereas the wet experiments showed a distributed pattern of shorter cracks. Therefore, the values that they obtained for $a_1$ and $a_2$ may be due to combining the results of the different types of experiments. If instead the wet and dry experiments had been analysed separately then, according to our theory, $a_2(\text{dry}) < a_2(\text{wet}) \leq 1.0$ due to the differing degree of localisation in the two types of experiments, whereas $a_1$ should equal 1.0 for both.

# 7 Summary and Conclusions

The two key assumptions of the published theory (Heffer & Bevan, 1990; Marrett & Allmendinger, 1991; Piggott, 1997; Berkowitz & Adler, 1998), are that (1) the spatial distribution of faults can be modelled as a Poisson process where fault positions are independently, uniformly distributed with a constant spatial fault intensity, and (2) that the size frequency distribution is a power law distribution. We find that, even when assumptions (1) and (2) are fulfilled, the difference of the estimated exponents from one and two dimensional samples of simulated fault patterns can deviate by 10–15% from the theoretically predicted difference for datasets consisting of 50–100 faults, using a maximum likelihood estimator for the exponent. For datasets consisting of $\sim 1000$ faults, the error is reduced to $\leq 5\%$. This deviation is due to the variance in the estimator which is used to calculate the exponent, and due to local fluctuations in fault density which arise naturally even for a spatial pattern that is statistically uniform. We further show that ambiguity concerning assumption (2), due to the effects of limited resolution and censoring on natural fault patterns, can lead to even larger deviations from the results theoretically predicted by previous workers. Specifically, we have studied the effect of fitting a power law exponent to data originating from an exponential distribution, to illustrate how an incorrect assumption of power law can lead to deviations from the results in Eq. (2) and (3). The deviation could be as large as 50%.

An important aspect of this paper concerns the effect of fault clustering on the relationship between fault populations sampled in one, two or three dimensions. In other words, we have studied in detail the impact of violating the assumption of spatial uniformity (assumption (1) above). In the first case we follow the approach of Malinverno (1997), and consider fault clustering perpendicular to a 1D traverse line which crosses a sub-parallel pattern of faults perpendicular to fault strike. If the distribution of fault centre points perpendicular to the traverse line is uniform, then $C_2 - C_1 = 1$ as previous theory predicts. However, when the fault centre points cluster in a self-similar fashion, the scaling parameter describing the spatial distribution modifies the relationship such that $C_2 - C_1 < 1$, depending on the degree of clustering. In the second case we consider the effect of fault clustering in a direction parallel to the 1D traverse line. This type of clustering is widely recognised in the field as en echelon arrays of fault segments which link as the faults grow during progressive deformation. We study the effect of en echelon clustering theoretically, and show how lack of information on linkage in 1D leads to problems when relating power law exponents obtained from 1D data to the corresponding exponent in 2D. The problem of relating 1D power law exponents to 2D exponents is that without any linkage criterion in 1D all faults are assumed to be isolated, while in 2D parameters such as the ratio of fault tip separation to overlap are often used to link fault segments together into longer structures.

To show how this affects the statistics we represent the overall fault population as a combination of isolated faults and en echelon fault arrays. En echelon arrays are those comprised of segments with overlap-to-separation ratios between their tips that have values within a specified interval (Aydin & Schultz, 1990; Huggins *et al.*, 1995; Gupta & Scholz, 1999). Separately, these two populations have different size-frequency distributions, with a higher proportion of small faults in the population of isolated faults. If the two separate populations are assumed to be power law distributions, represented by two different power law exponents, the combined distribution can be approximated by a third power law distribution. The combined distribution can be fitted by a power law exponent smaller than the exponent of the population of isolated faults, but larger than the exponent of the fault arrays. The relationship between these power law exponents is not found analytically, but is found via simulated fault patterns to depend on the ratio between the number of en echelon fault arrays and the total number of faults ($\theta$), and thus on the linkage criterion and the resolution of the data.

The appearance of the combined distribution is illustrated through simulation examples and theoretically predicted graphs (Figures 6, 7 and 8). For reasonably sized datasets, the deviations from a power law size-frequency distribution is not detectable visually. Using a range of values for $\theta$ from 0.25 – 0.5, we find that the approximate power law exponent of the combined distribution in 2D differs by 0.15 to 0.3 from that obtained when clustering and linkage are not taken into consideration. Consequently, we find that the difference between the 2D and 1D exponents is only 0.7-0.85 (i.e., $< 1.0$) for a fault population that contains en echelon clusters and obeys linear displacement-length scaling. If en echelon fault arrays dominate the fault pattern (i.e., $\theta > 0.5$) then the difference could be as low as 0.5 according to our simulations. These results show that if clustering and linkage of fault segments is ignored when analysing 1D data, predictions concerning the 2D and full 3D pattern, and/or the number of faults below the resolution of the data, will be meaningless.

# Appendix A: Length Distribution of a 1D Sample From a Generalised 2D Population – Assuming Spatial Uniformity

Marrett & Allmendinger (1991) showed the relationship between samples in different dimensions for the specific case of power law size-frequency distributions. Here we derive the analogous results for a general size-frequency distribution. For a given fault length distribution in 2D, it is of interest to determine the corresponding distribution for faults intersected by a 1D traverse. Let $(X, \Phi)$ be a pair of stochastic variables representing respectively the fault length in 2D and the angle between the fault strike and the traverse line. The pdf of $(X, \Phi)$ is denoted $f(x, \phi)$, and the event of a fault being intersected by a 1D traverse is denoted $I$. The pdf of lengths of faults intersected by the traverse differs from the marginal pdf $f(x)$, since small faults are under-sampled in lower dimensions. Let $f(x|i)$ be the pdf of fault lengths for the intersected faults, where we condition on the event $I$ of intersection. The pdf $f(x|i)$ is given by

$$f(x|i) = \frac{f(i,x)}{f(i)} = \frac{1}{f(i)} \int f(i|x,\phi) f(x,\phi)\, d\phi, \qquad (15)$$

where $f(i|x, \phi)$ is the probability of intersection conditioned on fault length and orientation and $f(i)$ is the unconditional probability of intersection. Under the assumption of independently, uniformly distributed fault positions, and fault sizes independent of the positions, the probability of a fault being intersected by the traverse is proportional to the fault length perpendicular to the direction of the traverse. This length is given as $X \sin \Phi$. Thus the probability of intersection, conditioned on fault length and orientation, is given as

$$f(i|x,\phi) = kx \sin\phi,$$

where $k$ is a normalisation constant. Assuming independence between fault length and orientation the joint pdf can be expressed as $f(x, \phi) = f(x)f(\phi)$. The numerator in Eq. (15) is then given as

$$f(i,x) = \int f(i|x,\phi) f(x) f(\phi)\, d\phi = kxf(x) \int \sin\phi\, f(\phi)\, d\phi = kxf(x) <\sin\phi>, \qquad (16)$$

and the unconditional probability of intersection is

$$f(i) = \int f(i,x)\, dx = k <\sin\phi> \int x f(x)\, dx = k <x><\sin\phi> . \qquad (17)$$

Substituting (16) and (17) in Eq. (15), we obtain

$$f(x|i) = \frac{kxf(x) <\sin\phi>}{k <x><\sin\phi>} = \frac{xf(x)}{<x>}. \qquad (18)$$

Thus the length distribution of intersected faults is independent of the distribution of fault orientation, under the assumption of independence between fault length and orientation. For the power law distribution in 2D, Eq. (1), we have

$$f(x) = \frac{\beta x_0^\beta}{x^{\beta+1}}, \qquad <x> = \frac{\beta x_0}{\beta - 1} \qquad \Rightarrow \qquad f(x|i) = \frac{(\beta-1)x_0^{\beta-1}}{x^\beta},$$

thus the length distribution in 1D is a power law distribution with exponent $\beta - 1$. The minimum fault length $x_0$ is the same both in 2D and 1D. For the exponential distribution in 2D, Eq. (7), we have

$$f(x) = \lambda \exp(-\lambda x), \qquad <x> = \lambda^{-1} \qquad \Rightarrow \qquad f(x|i) = \lambda^2 x \exp(-\lambda x),$$

thus the corresponding length distribution in 1D is a gamma distribution.

# Appendix B: Generalised Derivation of Slope on Bilogarithmic Plots of Fault Number Versus Size

Let $N(x)$ be the number of faults larger than $x$. This function is typically illustrated by plotting $\log N(x)$ versus $\log x$. The relative cumulative number $N(x)/N_T$, where $N_T$ is the total number of faults, gives an estimate of the cumulative distribution function $S(x)$ defined as

$$S(x) = \int_x^\infty f(x)\, dx$$

where $f(x)$ is the pdf of the length $X$. For a small number of observed faults, the uncertainty in the estimate $N(x)/N_T$ can be large, as illustrated in Section 1. However, for sufficiently large datasets, the slope of $N(x)$ on log-log plot is approximately equal to the slope of $\log S(x)$ plotted versus $\log x$. This slope is negative, with an absolute value given as

$$C(x) = -\frac{d \log S(x)}{d \log x} = -\frac{d \log S(x)}{dx} \cdot \frac{dx}{d \log x} = -\frac{x}{S(x)} \cdot \frac{dS(x)}{dx} = \frac{x f(x)}{S(x)}.$$

For the power law distribution, Eq. (1), the function $S(x)$ and the slope $C(x)$ are given as

$$S(x) = \left(\frac{x_0}{x}\right)^\beta$$
$$C(x) = x \cdot \frac{\beta x_0^\beta}{x^{\beta+1}} \cdot \left(\frac{x_0}{x}\right)^{-\beta} = \beta,$$

where $x_0$ is the minimum fault size. Thus the slope $C(x) = \beta$ is constant for all values of $x$. For the exponential distribution with intensity $\lambda = <x>^{-1}$, Eq. (7), the function $S(x)$ and the absolute value $C(x)$ of the slope are given as

$$S(x) = \exp(-\lambda x)$$
$$C(x) = x \cdot \lambda \exp(-\lambda x) \cdot \exp(\lambda x) = \lambda x,$$

and for the gamma distribution, Eq. (9), the following results are obtained:

$$S(x) = (\lambda x + 1)\exp(-\lambda x)$$
$$C(x) = x \cdot \lambda^2 x \exp(-\lambda x) \cdot \frac{\exp(\lambda x)}{\lambda x + 1} = \frac{(\lambda x)^2}{\lambda x + 1}.$$

# Appendix C: Random Linkage Criterion for Faults in 1D

The overlap-to-separation ratio, $O/S$ (Fig. 4), can be used as a linkage criterion for faults in 2D, as discussed in Section 5. However, if there is only 1D data available the separation $S$ is measured, but not the overlap $O$. Moreover, only the displacements and not the lengths can be measured along a 1D traverse line. Using the relationship $D_{\max} = \gamma L^n$ between length and maximum displacement, and a model for the displacement profile (e.g., Eq. (4)), information from the observed displacements along a traverse can be used to constrain the lengths and thus decide if two faults closely located on the traverse may be classified as linked. The $O/S$ ratio can still not be found exactly, but a statistical approach can be used in the linkage classification. A random linking of faults can be applied, based on the probability of the $O/S$ ratio falling in the specified interval; $\text{Prob}\{M_1 \leq O/S \leq M_2\}$. This probability can then be conditioned on the observed separation; $\text{Prob}\{M_1 \leq O/S \leq M_2 | S = s\}$.

Consider two neighbouring faults with the same dip and dip direction along the traverse, with lengths following a power law distribution with exponent $b$ and lower limit of fault length $l_0$. Assume the intersection points with the traverse are uniformly distributed along the fault traces. The distribution of the overlap $O$ is then found to be

$$f(o) = \begin{cases} \left(\dfrac{b}{b+1}\right)^2 \dfrac{2o}{l_0^2}, & 0 \leq o \leq l_0 \\ \dfrac{2b+1}{(b+1)^2} \cdot \dfrac{2b l_0^{2b}}{o^{2b+1}}, & o \geq l_0 \end{cases} \tag{19}$$

The distribution of the overlap has a power law tail with exponent $2b$ for overlaps larger than the minimum fault length $l_0$. The conditional probability of $M_1 \leq O/S \leq M_2$ given $S = s$ can be reformulated as $\text{Prob}\{M_1 \leq O/S \leq M_2 | S = s\} = \text{Prob}\{sM_1 \leq O \leq sM_2\}$, which is found to be

$$\mathrm{Prob}\{sM_1 \le O \le sM_2\} \;=\; \int_{sM_1}^{sM_2} f(o)\,do \;=$$

$$\begin{cases} \left(\dfrac{b}{b+1}\right)^2 \left(\dfrac{s}{l_0}\right)^2 (M_2^2 - M_1^2), & 0 \le s \le \dfrac{l_0}{M_2} \\ 1 - \left(\dfrac{b}{b+1}\right)^2 \left(\dfrac{s}{l_0}\right)^2 M_1^2 - \dfrac{2b+1}{(b+1)^2} \left(\dfrac{s}{l_0}\right)^{-2b} M_2^{-2b}, & \dfrac{l_0}{M_2} \le s \le \dfrac{l_0}{M_1} \\ \dfrac{2b+1}{(b+1)^2} \left(\dfrac{s}{l_0}\right)^{-2b} (M_1^{-2b} - M_2^{-2b}), & s \ge \dfrac{l_0}{M_1} \end{cases} \tag{20}$$

Hence, if all parameters $M_1$, $M_2$, $b$ and $l_0$ are known, or estimated, the probability $\mathrm{Prob}\{sM_1 \le O \le sM_2\}$ can be found for a given separation $S = s$. The power law exponent of displacements observed along the traverse is $C_1' = \beta_S - 1/n$, see Section 5. The power law exponent of the corresponding lengths is then given by $b = C_1 = nC_1' = n\beta_S - 1$. Thus if $n$ is known (e.g., $n = 1$), the exponent $b$ can be found based on the observations. If the constant $\gamma$ in the expression $D_{\max} = \gamma L^n$ is known, the lower limit $l_0$ of fault length can be obtained from a lower limit $d_0$ of displacements along the traverse.

A random linking criterion can be chosen as "Classify two faults with the same dip and dip direction, separated by a distance $s$, as linked with probability $\mathrm{Prob}\{sM_1 \le O \le sM_2\}$". This can be done by generating a random number $u$ from the uniform distribution on the unit interval $[0,1]$. If $u \le \mathrm{Prob}\{sM_1 \le O \le sM_2\}$, which occurs with probability $\mathrm{Prob}\{sM_1 \le O \le sM_2\}$, the faults are classified as linked. If $u > \mathrm{Prob}\{sM_1 \le O \le sM_2\}$, the two observations are treated as isolated faults. Note that this procedure does not give a unique linkage of the faults, only one statistical realisation.

The probability $\mathrm{Prob}\{M_1 \le O/S \le M_2\}$ can not be related directly to the parameter $\theta$ given in Eq. (13). However, based on 2D observations, $\mathrm{Prob}\{M_1 \le O/S \le M_2\}$ can empirically be found as

$$\mathrm{Prob}\{M_1 \le O/S \le M_2\} = \frac{N_S}{N_I + N_S} \tag{21}$$

where $N_I$, as before, is the number of isolated faults and $N_S$ is the total number of single fault segments included in linked fault structures. Each linked fault system contains several single fault segments, thus $N_S > N_A$, resulting in $\mathrm{Prob}\{M_1 \le O/S \le M_2\} > \theta$.

## Acknowledgements

We are grateful to Alberto Malinverno for introducing us to his work on turbidite beds, which Section 3 is based on. We thank Chris Townsend (Statoil R&D) for instigating and supporting this collaborative work. We thank Chris Scholz, Ian Main and Henning Omre for reviews and discussions. H. Borgos is funded by a PhD grant from VISTA (*Det Norske Videnskaps-Akademi* and *Statoil*). P. Cowie is supported by a Research Fellowship from The Royal Society of London. N. Dawers was funded by NERC-ROPA grant 95/21(GR3/RO43) to J. Underhill and by the Edinburgh Department of Geology and Geophysics.

## References

Ackermann, R. V., Schlische, R. W. & Withjack, M. O. (1997). Systematics of an evolving population of normal faults in scaled physical models, *Abstracts with Programs*, Vol. 29, Geological Society of America, p. 198.

Aydin, A. & Schultz, R. A. (1990). Effect of mechanical interaction on the developement of strike-slip faults with echelon patterns, *Journal of Structural Geology* **12**(1): 123–129.

Berkowitz, B. & Adler, P. M. (1998). Stereological analysis of fracture network structure in geological formations, *Journal of Geophysical Research* **103**(B7): 15339–15360.

Bour, O. & Davy, P. (1999). Clustering and size distributions of fault patterns: Theory and measurements, *Geophysical Research Letters* **26**(13): 2001–2004.

Clark, R. M., Cox, S. J. D. & Laslett, G. M. (1999). Generalization of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats, *Geophysical Journal International* **136**: 357–372.

Cowie, P. A. (1998). Normal Fault Growth in Three-Dimensions in Continental and Oceanic Crust, *Faulting and Magmatism at Mid-Ocean Ridges*, Geophysical Monograph 106, the American Geophysical Union, pp. 325–348.

Cowie, P. A. & Scholz, C. H. (1992a). Physical explanation for the displacement-length relationship of faults using a post-yield fracture mechanics model, *Journal of Structural Geology* **14**(10): 1133–1148.

Cowie, P. A. & Scholz, C. H. (1992b). Displacement-length scaling relationship for faults: data synthesis and discussion, *Journal of Structural Geology* **14**(10): 1149–1156.

Cowie, P. A., Malinverno, A., Ryan, W. B. F. & Edwards, M. H. (1994). Quantitative fault studies on the East Pacific Rise: A comparison of sonar image techniques, *Journal of Geophysical Research* **99**(B8): 15205–15218.

Dawers, N. H. (1996). Fault size distribution on the Volcanic Tableland, California: implications for small-scale fault prediction, *Extended Abstract Volume: Faulting, Fault Sealing and Fluid Flow in Hydrocarbon Reservoirs*, University of Leeds Rock Deformation Research Group, in conjunction with the Petroleum Group & Tectonic Studies Group, The Geological Society, London, pp. 88–89.

Dawers, N. H., Anders, M. H. & Scholz, C. H. (1993). Growth of normal faults: Displacement-length scaling, *Geology* **21**: 1107–1110.

Dawers, N. H. & Anders, M. H. (1995). Displacement-length scaling and fault linkage, *Journal of Structural Geology* **17**(5): 607–614.

Hatton, C. G., Main, I. G. & Meredith, P. G. (1993). A comparison of seismic and structural measurements of scaling exponents during tensile subcritical crack-growth, *Journal of Structural Geology* **15**(12): 1485–1495.

Heffer, K. J. & Bevan, T. G. (1990). Scaling Relationships in Natural Fractures: Data, Theory, and Application, *Europec 90*, Society of Petroleum Engineers, pp. 367–376. SPE Paper 20981.

Huggins, P., Watterson, J., Walsh, J. J. & Childs, C. (1995). Relay zone geometry and displacement transfer between normal faults recorded in coal-mine plans, *Journal of Structural Geology* **17**(12): 1741–1755.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2 edn, Wiley.

Malinverno, A. (1997). On the power law size distribution of turbidite beds, *Basin Research* **9**: 263–274.

Marrett, R. (1996). Aggregate properties of fracture populations, *Journal of Structural Geology* **18**(2/3): 169–178.

Marrett, R. & Allmendinger, R. W. (1991). Estimates of strain due to brittle faulting: sampling of fault populations, *Journal of Structural Geology* **13**(6): 735–738.

Marrett, R. & Allmendinger, R. W. (1992). Amount of extension on "small" faults: An example from the Viking graben, *Geology* **20**: 47–50.

Peacock, D. C. P. & Sanderson, D. J. (1991). Displacements, segment linkage and relay ramps in normal-fault zones, *Journal of Structural Geology* **13**(6): 721–733.

Pickering, G., Bull, J. M. & Sanderson, D. J. (1995). Sampling power-law distributions, *Tectonophysics* **248**: 1–20.

Pickering, G., Bull, J. M. & Sanderson, D. J. (1996). Scaling of fault displacements and implications for the estimation of sub-seismic strain, *in* P. C. Buchanan & D. A. Nieuwland (eds), *Modern Developements in Structural Interpretation, Validation and Modeling*, Geological Society Special Publication No. 99, pp. 11–26.

Piggott, A. R. (1997). Fractal relations for the diameter and trace length of disc-shaped fractures, *Journal of Geophysical Research* **102**(B8): 18121–18125.

Schlische, R. W., Young, S. S., Ackermann, R. V. & Gupta, A. (1996). Geometry and scaling relations of a population of very small rift-related normal faults, *Geology* **24**(8): 683–686.

Scholz, C. H., Dawers, N. H., Yu, J. Z. & Anders, M. H. (1993). Fault growth and fault scaling laws – preliminary results, *Journal of Geophysical Research* **98**(B12): 21951–21961.

Schultz, R. (in press). Fault-population statistics at the valles marineris extensional province, mars: implications for segment linkage, crustal strains, and its geodynamical development, *Tectonophysics*.

Segall, P. & Pollard, D. D. (1983). Joint formation in granitic rocks of the sierra nevada, *The Geological Society of America Bulletin* **94**: 563–575.

Spryopoulos, C., Griffith, W. J., Scholz, C. H. & Shaw, B. E. (1999). Experimental evidence for different strain regimes of crack populations in a clay model, *Geophysical Research Letters* **26**(8): 1081–1084.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2 edn, Wiley, New York.

Walsh, J. J., Watterson, J. & Yielding, G. (1991). The importance of small-scale faulting in regional extension, *Nature* **351**: 391–393.

Watterson, J. (1986). Fault dimensions, displacement and growth, *Pure and Applied Geophysics* **124**: 365–373.

Willemse, E. J. M., Pollard, D. D. & Aydin, A. (1996). Three-dimensional analyses of slip distributions on normal fault arrays with consequences for fault scaling, *Journal of Structural Geology* **18**(2/3): 295–309.

Wojtal, S. F. (1996). Changes in fault displacement populations correlated to linkage between faults, *Journal of Structural Geology* **18**(2/3): 265–279.