

1 Supplementary file

2 Distribution fitting procedure

3 Clauset et al. (2009) and Rizzo et al (2017) proposed that the maximum likelihood estimator (MLE)
4 should be preferred over use of least square regression analyses (R^2) for the fitting of power-law
5 distributions. Rizzo et al. (2017) performed the MLE on power-law, log-normal and exponential
6 distributions by using a suite of custom MATLAB™ functions, integrated into FracPaQ (Healy et al.,
7 2017). The MLE approach maximizes the likelihood, gives estimate of the governing parameters (α
8 for power-law distribution, λ for exponential distribution and μ and σ for the log-normal distribution)
9 of the different fitting equations:

$$\text{Power-law:} \quad p(x|\alpha) = \frac{\alpha-1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \quad \text{Eq. 1}$$

$$\text{Log-normal:} \quad p(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad \text{Eq. 2}$$

$$\text{Exponential:} \quad p(x|\lambda) = \lambda \exp(-\lambda x) \quad \text{Eq. 3}$$

10

11 where x_{\min} in the power-law distribution, is a required parameter representing the lower bound
12 below which the power-law distribution is not valid (Clauset et al., 2009). The x_{\min} parameter can be
13 estimated using the Kolmogorov-Smirnoff (KS) test which minimizes the difference between the data
14 and the synthetic data generated using the parameters derived from the MLE (Clauset et al., 2009;
15 Rizzo et al., 2017). The K-S test generates an H-percentage (HP) which is the probability of accepting
16 the H_0 (null hypothesis) result over the total n -cycles (in this case $n = 2500$). If the p -value is less
17 than or equal to 0.05, the test suggests that “the observed data are inconsistent with the null
18 hypothesis, so the null hypothesis must be rejected, while if the p -value is far from zero and close to
19 1, the observed data are not inconsistent with the null hypothesis, and the chosen fitting method
20 can be applied” (Light et al., 2009). However having a p -value larger than 0.05, does not prove that
21 the tested hypothesis is the most appropriate distribution. Clauset et al. (2009) suggest that the p -
22 value for alternative distributions can be calculated to test between other possible distributions. We

23 used the KS probability value (reported in tables S2 & S2) to decide between log normal and power
24 law distributions. In almost all examples, the exponential fit yielded a noticeably poorer result so is
25 discounted.

26

27 Knowing that attributes collected from outcrop are naturally affected by truncation and
28 censoring bias, we also performed the MLE and KS test for truncated populations defined by
29 step-wise removal of data values from either end of the distribution (Fig. S1). We term these
30 the upper cut (uc) and lower cut (lc) respectively. 40 values of censoring and truncation for
31 uc and lc were considered, resulting in 800 simulations. The resulting values of KS values were
32 visualized a checkerboard-like plots (e.g. Fig S1). The best-fit results produce the highest
33 percentage values red colours in Fig 5.

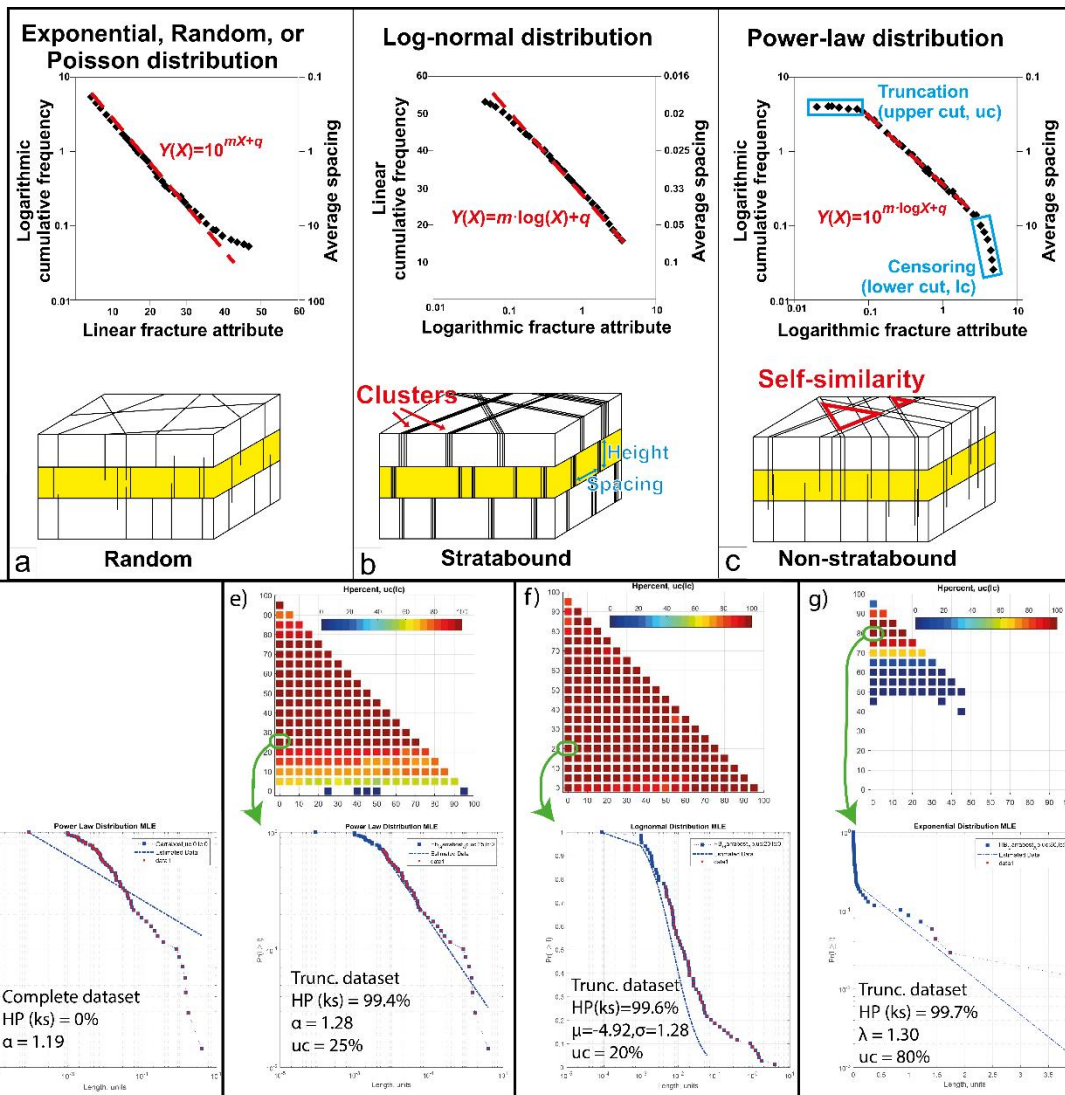
34

35 In Table S1 and S2 we report the KS value (HP), and distribution coefficients obtained using
36 the MLE for both non-truncated (power-law distributions) and for truncated (power-law
37 distribution, exponential and log normal) populations aperture and length.

38

39 Complete (non-truncated) populations are generally best described by a log-normal
40 distribution (they show consistently high percentage fitting values). However, the KS values
41 for truncated log-normal and truncated power-law datasets are similar suggesting either
42 distribution might be preferred. The choice of the best-fit distribution should not be based on
43 complete population because population “end-points” (blue regions in Fig S1) are biased. For
44 the datasets the aperture data can mostly be well describe by power-law distribution whereas
45 the length data could be described by either power-law or log normal distributions.

46



47

48 Fig S1. Example fracture distributions and models that show their typical development (after Dichiarante et al In
 49 Review) a) An exponential or random distribution, b) a log-normal distribution producing fracture corridors and
 50 c) a scale invariant power law distribution. d) an example of the modified Rizzo et al (2017) MLE distribution
 51 fitting procedure for a basement outcrop dataset from Garrabost, Hebrides. Complete dataset shows a poor
 52 power law fit, e) same dataset with a truncated (25%) picked as the minimum value of uc/lc from checkerboard
 53 of KS results that gives a good fit (>99% Ks value – dark red colour). f) the log-normal result for same dataset
 54 and g) the exponential result. All of these fits are acceptable, however the exponential can be dismissed as
 55 80% of the data have been removed from the analysis, the power law and log normal are very close with a slight
 56 preference for the log normal distribution (slightly higher KS value and slightly more of the data included).