Supplementary Data File

Fault and fracture scaling and connectivity in the Devonian Orcadian Basin and implications for the offshore Clair Field, Scotland

A.M. DICHIARANTE, K.J.W. McCAFFREY, R.E. HOLDSWORTH, T.I. BJØRNARÅ and E.D. DEMPSEY

Testing Fracture Distributions

Mitzenmacher (2004) showed that difference in the tail of the power-law, log-normal and double Pareto (a distribution with the body of a log-normal and tail of power-law distribution) are may be better visualized in complementary cumulative distribution plots at logarithmic scale (ccdf, **Fig. S1**). Legitimate data points (data occurring in the yellow box in **Fig. S1**) are well fitted by power-law, log-normal and double Pareto distribution (Mitzenmacher, 2004). However, when data are affected by censoring and truncation effects, the power-law distribution type shows the least similarity to the biased data. More specifically, when datasets include truncated data (blue box in **Fig. S1**), both log-normal and double Pareto distributions fit the data (see Truncation blue box in **Fig. S1**a) and therefore it is not correct to base the distribution choice on these biased data (Mitzenmacher, 2004).

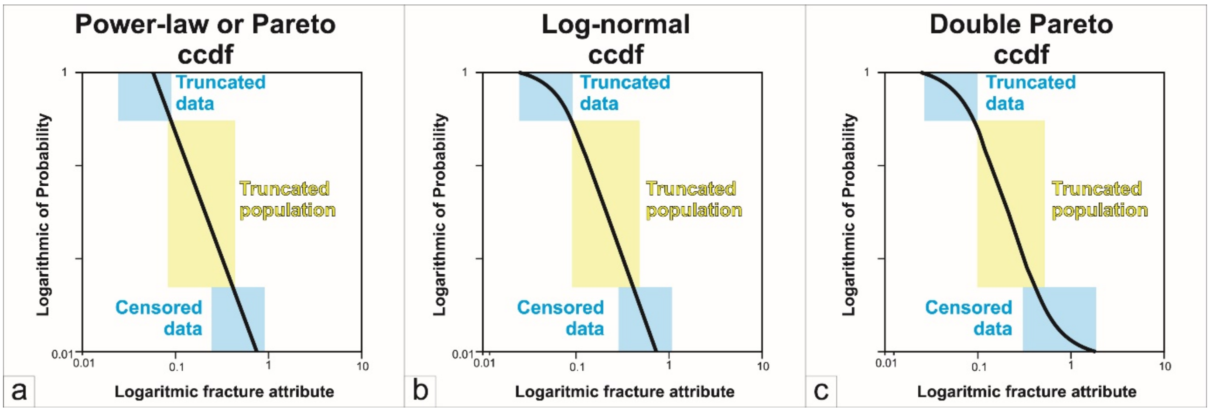

Fig S1: Sketch showing the similarity in the shape of (a) power-law, (b) log-normal and (c) double pareto complementary cumulative distribution function (ccdf) as illustrated in Mitzenmacher (2004). The yellow region represents the truncated part of the population that do not suffer of truncation and censoring biases, blue regions represent truncated and censored data.

1.1.1 Maximum likelihood estimator (MLE) and Kolmogorov-Smirnoff (KS)

The determination of which type of population is most likely to have generated the sample recorded is a key part of any fracture attribute analysis (Rizzo et al., 2017). As pointed out by Clauset et al. (2009), use of the maximum likelihood estimator (MLE) should be preferred over use of least square regression analyses (R^2) for the fitting of power-law distributions because a power-law may appear to be a good fit even when the data are non-power-law. Rizzo et al. (2017) performed the MLE on power-law, log-normal and exponential distributions by using a suite of custom MATLAB™ functions, integrated into FracPaQ (Healy et al., 2017). They compared the MLE to the linear regression method for synthetic data in order to demonstrate the validity and ability of this approach to correctly estimate statistical parameters. The MLE approach maximizes the likelihood, gives estimate of the governing parameters (α for power-law distribution, λ for exponential distribution and μ and σ for the log-normal distribution) of the different fitting equations:

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

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

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

where $x_{\min}$ in the power-law distribution, is a required parameter representing the lower bound below which the power-law distribution is not valid (Clauset et al., 2009). The $x_{\min}$ parameter can be estimated using the Kolmogorov-Smirnoff (KS) test which minimizes the difference between the data and the synthetic data generated using the parameter derived from the MLE (Clauset et al., 2009; Rizzo et al., 2017). Two percentage outputs are obtained from the MLE method. The P-percentage (PP) and H-percentage (HP) are the percentages of the *p*-value larger than 0.05 over the total *n*-cycles and the percentage of the H0 (null hypothesis) result over the total *n*-cycles, respectively. If the *p*-value is less than or equal to 0.05, the test suggests that "the observed data are inconsistent with the null hypothesis, so the null hypothesis must be rejected, while if the *p*-value is far from zero and close to 1, the observed data are not inconsistent with the null hypothesis, and the chosen fitting method can be applied" (Hung et al., 1997). However having a *p*-value larger than 0.05, does not prove that the tested hypothesis is true. Clauset et al. (2009) have shown that the *p*-value for alternative distributions can be calculated to test against other possibilities.

When testing the entire sample we might obtain misleading fitting results because we have included censored and truncated data. On the other hand, when testing truncated populations we could be removing legitimate points and increasing the error; for example, if the upper cut is too high (x_{min} too high) (Clauset et al., 2009). In order to address this problem, the methodology proposed by Rizzo et al. (2017) and used by FracPaQ (Healy et al., 2017), was implemented here by calculating the MLE on progressively truncated populations for power-law, exponential and log-normal distributions. Knowing that attributes collected from outcrop are naturally affected by truncation and censoring bias, we performed the MLE and KS test and calculated the PP and HP for truncated populations defined by progressive variation in the upper cut (uc) and lower cut (lc). 40 values of censoring for both uc and lc were considered, resulting in 800 simulations. The resulting values of percentages (HP and PP) were visualized in two lc vs. uc checkerboard-like plots (e.g. **Fig. S2**). The best-fit results in the highest percentage values obtained, with the minimum lower and upper cuts (red colours in **Fig. S2** and corresponding distribution plot in **Fig. S3**). This methodology is more robust than the very commonly used visual estimation (**Fig. S2**). Such an analysis shows that the upper cut (lower values) has a consistently greater influence over the population fit compared to the lower cut (see **Fig. S2** and percentages reported in Tables 3 and 4).

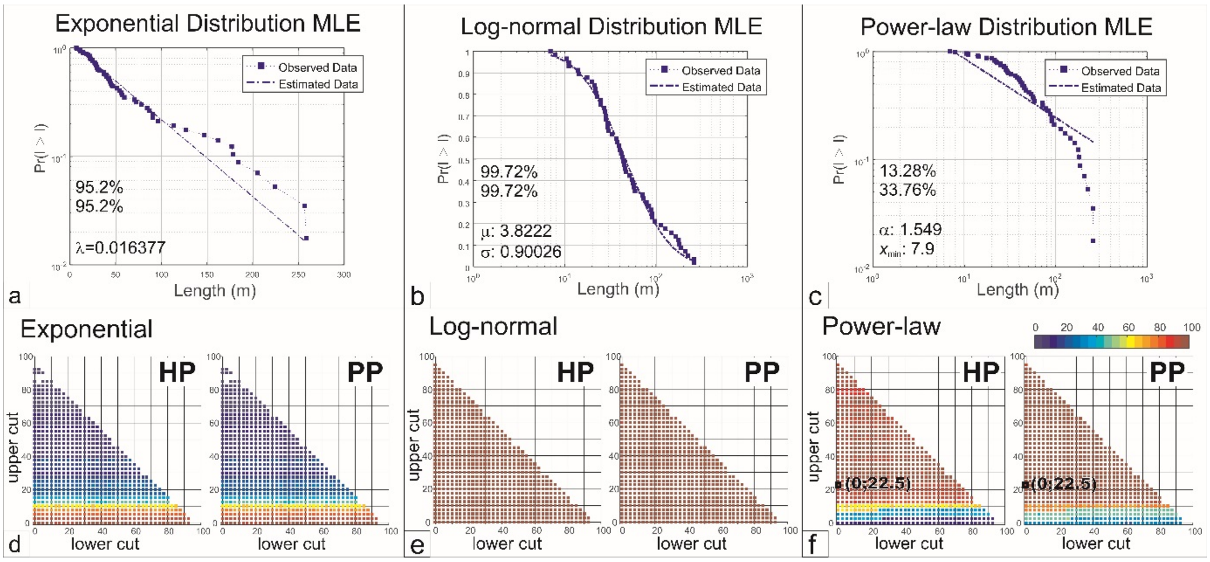

Fig. S2: (Top) Example of MLE for (a) exponential, (b) log-normal and (c) power-law distributions for the entire length population; (Bottom) "Checkerboard" diagrams showing the values of HP and PP for (d) exponential, (e) log-normal and (f) power-law distributions of progressively truncated population. Data are from the transect at St. John's Point (SJ) in this study.

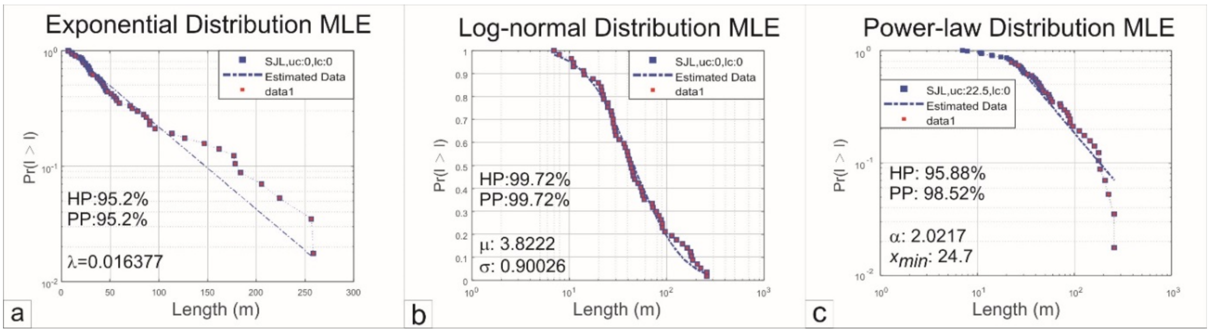

Fig. S3: MLE plot of the highest value of HP and PP for (a) exponential distribution, (b) log-normal distribution and (c) power-law distributions obtained with the minimum truncation (lc = 0; uc = 0 for exponential and log-normal and lc = 0; uc = 22.5) in the population. Dashed blue line is the fitting curve and red squares are the truncated data used for the fitting. Data are from the transect at St. John's Point (SJ) in this study.

Tab. 1: H-percentage, P-percentage and coefficients obtained by using the MLE in both un-truncated and truncated power-law population for length. EX = exponential, LN = log-normal, PL = power-law.

Length											
			BTr1	BTr2	CTr1	CTr2	TTr1	TTr2	SK	DO	SJ
Entire population	**EX**	PP	82	12.5	97.9	98.3	99.7	92.4	97.2	91.7	95.2
		HP	82	12.5	97.9	98.3	99.7	92.4	97.2	91.7	95.2
		λ:	1.3	1.1	0.4	0.6	2.4	3.2	73474.28	0.0165	0.0163
	LN	PP	95.8	98.9	97.7	99.7	99.6	96.7	99.7	98.6	99.7
		HP	95.8	89.9	97.7	99.7	99.6	96.7	99.7	98.6	99.7
		μ:	-0.9	-1.1	0.2	0.1	-1.3	-1.5	-11.42	3.7	3.8
		σ:	1.225	1.303	1.27	0.929	1.038	1.036	0.851	0.901	0.900
	PL	PP:	0	0	49.8	2.6	50.2	16.5	58.2	0.12	13.3
		HP:	0	0	64	12.6	77.5	30.6	74.6	0.56	33.8
		α:	1.3	1.2	1.4	1.49	1.5	1.4	1.5	1.4	1.5
		x_{min}	0.01	0.005	0.155	0.115	0.05	0.023	$1.5e^{-06}$	3.5	7.9

Truncated Population	**PL**	uc:	-	0	0	0	2.5	-	0	0	0
		lc:	-	30	30	30	45	-	10	15	12
		PP:		95.5	61	89.4	87.7		92.5	84.2	91.6
		HP:		97.8	77.1	94.4	94.8		97.6	95.2	97.6
		α:		1.840	1.781	2.156	2.152		1.793	1.994	1.923
		x_{min}		0.185	0.88	0.80	0.27		$3.8e^{-06}$	20.5	19.9

Tab. 2: H-percentage, P-percentage and coefficients obtained by using the MLE in both un-trucated and power-law truncated population for aperture. EX = exponential, LN = log-normal, PL = power-law.

Aperture									
			BTr1	BTr2	CTr1	CTr2	TTr1	TTr2	SK
Entire population	**EX**	PP:	54.8%	72.6%	81.9%	16.4%	99.8%	81.2%	97.8%
		HP:	54.8%	72.6%	81.9%	16.4%	99.8%	81.2%	97.8%
		λ:	197.461	482.655	341.705	263.194	1055.5947	328.879	73474.286
	LN	PP:	99.6%	99.6%	99.9%	99.3%	99.4%	99.6%	99.8%
		HP:	99.6%	99.6%	99.9%	99.3%	99.4%	99.6%	99.8%
		μ:	-6.803	-7.608	-7.289	-7.395	-7.743	-6.840	-11.42
		σ:	2	1.773	2.043	1.980	1.507	1.3055	0.851
	PL	PP:	51.7%	69.9%	82.5%	78.5%	71.1%	95.8%	52.9%
		HP:	71.2%	86.6%	82.5%	78.5%	88.1%	98.9%	74.6%
		α:	1.202	1.240	1.252	1.231	1.248	1.516	1.489
		x_{min}	$1e^{-05}$	$1e^{-05}$	$2e^{-05}$	$1e^{-05}$	$1e^{-05}$	0.0002	$1.5555e^{06}$
Truncated Population	**PL**	uc:	0	0	0	0	0	0	0
		lc:	15	15	22.5	20	27.5	0	15
		PP:	93.2%	97.4%	97.3%	95.2	97.6%	95.8%	96.4%
		HP:	98.1%	97.4%	97.3%	98.8	97.6%	98.9%	96.4%
		α:	1.414	1.454	1.4343	1.3964	1.8177	1.516	1.8781
		x_{min}	0.0002	0.0001	0.0002	0.0001	0.0003	0.0002	$4.66e^{-06}$

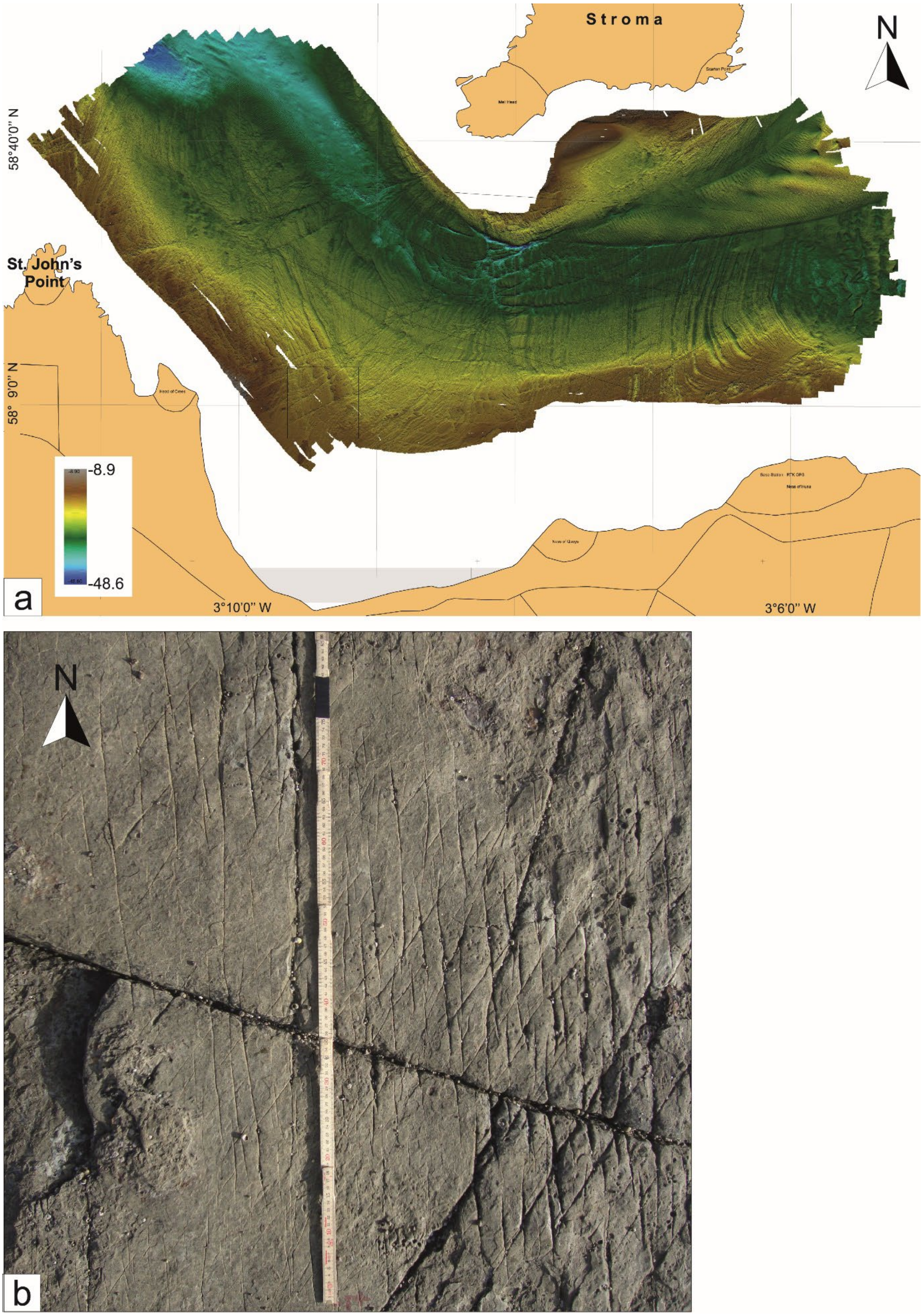

Fig. S3: Raw images used for 2D analysis: (a) Bathymetric data from the area between St. John's Point and Stroma Island and (b) outcrop pavement photograph.

Clauset, A., Shalizi, C. R. and Newman, M. E.: Power-law distributions in empirical data. SIAM review, v. 51, no. 4, p. 661-703, 2009.

Hung, H.M.J., O'Neill, R.T., Bauer, P., Kohne, K.: The behavior of the p-value when the alternative hypothesis is true. Biometrics. 53 (1): 11–22, 1997.

Healy, D., Rizzo, R. E., Cornwell, D. G., Farrell, N. J., Watkins, H., Timms, N. E., Gomez-Rivas, E. and Smith, M.: FracPaQ: A MATLAB™ toolbox for the quantification of fracture patterns. Journal of Structural Geology, v. 95, p. 1-16, 2017.

Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions. Internet mathematics, 1(2), 226-251, 2004.

Rizzo R. E., Healy, D. and De Siena, L.: Benefits of maximum likelihood estimators for fracture attribute analysis: Implications for permeability and up-scaling. Journal of Structural Geology, v. 95, p. 17-31, 2017.