Pure appl. geophys. 158 (2001) 741–757
0033–4553/01/040741–17 $ 1.50 + 0.20/0



Pure and Applied Geophysics

# Upper-truncated Power Laws in Natural Systems

STEPHEN M. BURROUGHS[1] and SARAH F. TEBBENS[2]

*Abstract*—When a cumulative number-size distribution of data follows a power law, the data set is often considered fractal since both power laws and fractals are scale invariant. Cumulative number-size distributions for data sets of many natural phenomena exhibit a "fall-off" from a power law as the measured object size increases. We demonstrate that this fall-off is expected when a cumulative data set is truncated at large object size. We provide a generalized equation, herein called the General Fitting Function (GFF), that describes an upper-truncated cumulative number-size distribution based on a power law. Fitting the GFF to a cumulative number-size distribution yields the coefficient and exponent of the underlying power law and a parameter that characterizes the upper truncation. Possible causes of upper truncation include data sampling limitations (spatial or temporal) and changes in the physics controlling the object sizes. We use the GFF method to analyze four natural systems that have been studied by other approaches: forest fire area in the Australian Capital Territory; fault offsets in the Vernejoul coal field; hydrocarbon volumes in the Frio Strand Plain exploration play; and fault lengths on Venus. We demonstrate that a traditional approach of fitting a power law directly to the cumulative number-size distribution estimates too negative an exponent for the power law and overestimates the fractal dimension of the data set. The four systems we consider are well fit by the GFF method, suggesting they have properties characterized by upper-truncated power laws.

**Key words:** Truncated power law, fault length and offset, hydrocarbon volume, forest fire area, size distribution, fractals.

## 1. Introduction

The power law has been fit to number-size data distributions of numerous natural phenomena, from the Gutenberg-Richter number-size distribution of earthquake magnitude (GUTENBERG and RICHTER, 1949) to the cumulative number-size distribution of cloud areas observed in satellite and space shuttle cloud images (PELLETIER, 1997). Once identified, the power law may be used as the basis for probabilistic prediction of data behavior (MALAMUD *et al.*, 1996). The power law has been the basis for modeling many processes, such as rock fragmentation (TURCOTTE, 1997), forest fire areas (MALAMUD *et al.*, 1998), landslide areas (PELLETIER *et al.*,

[1] Department of Marine Science, University of South Florida, St. Petersburg, Florida 33701, U.S.A.; also at Department of Chemistry and Physics, The University of Tampa, Tampa, Florida 33606, U.S.A.
[2] Department of Marine Science, University of South Florida, St. Petersburg, Florida 33701, U.S.A.

1997), and hydrocarbon volumes (BARTON and SCHOLZ, 1995). A fundamental aspect of such studies is fitting a power law to the observed cumulative and/or non-cumulative data.

When examining log–log plots of cumulative number versus object size, there are often three distinct regions on the graph. At the smallest object sizes the slope of the graph approaches zero. This is understood because the smallest objects are most easily missed, resulting in an incomplete data set. Below the resolution of size measurement no objects are observed, therefore the slope of the cumulative distribution is zero. In the middle of the data range the graph often approximates a straight line with a negative slope. This is the expected behavior of a negative power law on a log–log graph. Many authors have fit a power law to this straight portion of the graph (e.g., JACKSON and SANDERSON, 1992; LU *et al.*, 1999; SCHOLZ, 1998; SORNETTE *et al.*, 1993). At the largest plotted object sizes, the data points often fall below the power-law trend of the straight portion of the graph (e.g., BARTON and SCHOLZ, 1995; MALINVERNO, 1997; MARITAN *et al.*, 1996; WANG, 1995). In the "fall-off" region there are fewer large objects than predicted by the power-law trend. In some cases, authors fit a power law to only these largest object sizes (VILLEMIN *et al.*, 1995). This fall-off has been noticed for models of both synthetic and natural data (e.g., BROOKS and ALLMENDINGER, 1996; LA BARBERA and ROTH, 1994; MALAMUD *et al.*, 1998; NAGATANI, 1993; PICKERING *et al.*, 1995, 1994; WANG, 1995), as discussed in Section 7.1. Equations derived from these models are specific to the system under consideration.

The approach of this work is more general, and the fitting function for an upper-truncated power law appears to have broad application. We start with the general properties of number-size distributions that follow power laws (Section 2) and upper-truncate these distributions (Section 3). We provide a single function that describes upper-truncated distributions, herein called the General Fitting Function (GFF) (Section 4). The GFF fits both the straight region and the fall-off region on a log–log graph of the distribution. The fall-off is a predictable characteristic of a power-law distribution that is upper-truncated (Section 5). The GFF is found to fit the cumulative number-size distribution of measurements characterizing four natural systems (Section 6).

## 2. Non-cumulative and Cumulative Number-size Distributions

Non-cumulative and cumulative number-size functions are the basis of all equations discussed in this work and are defined below. In this work, the terms *power function* and *power law* are equivalent and refer to continuous functions, whereas *distribution* is used to describe a set of discrete objects. *Objects* are the items measured to construct a data set, such as earthquakes, faults, and forest fires. Each measured object has a characteristic size such as earthquake *magnitude*, fault *offset length*, or

forest fire *area*. Non-cumulative distributions include only the number of objects within each data bin. These distributions also are referred to as density distributions and may be plotted as histograms. *Cumulative* distributions include the number of objects within each data bin plus all larger objects. Cumulative number-size distributions sometimes are plotted as rank-order distributions or Zipf plots (MANDELBROT, 1997 and references therein).

### *2.1 Non-cumulative Number-size Functions*

For a set of data points in which the number-size distribution follows a power law, the number of objects, $n(r)$, with characteristic size $r$, is

$$n(r) = cr^{-d}. \tag{1}$$

On a log–log plot of $n(r)$ versus $r$, this equation is a straight line where $-d$ equals the slope and $c$ is a constant equal to the number of objects with size $r = 1$.

### *2.2 Cumulative Number-size Functions*

Cumulative number-size functions are commonly used in fractal analysis of data sets to find the fragmentation dimension. The number-size distribution for a large number of objects may be fractal if the cumulative number of objects, $N(r)$, with characteristic size greater than or equal to $r$ satisfies the power-law relation

$$N(r) = Cr^{-D}. \tag{2}$$

Equation (2) is a straight line on a log–log plot, where $-D$ is the slope and $C$ is a constant equal to the number of objects with size $r \geq 1$. This cumulative relation applies when $r$ is a continuous set of values. When the characteristic size, $r$, is linear, $D$ is the fractal dimension of the distribution. If the characteristic size is expressed as an area or a volume, the fractal dimension is $2D$ or $3D$, respectively. Equations (1) and (2) are consistent with the work of many authors (e.g., FEDER, 1988; TURCOTTE, 1997).

### *2.3 The Effect of Binning*

The manner of binning object sizes is an important consideration, as binning affects the relationship between non-cumulative (equation 1) and cumulative (equation 2) distributions (MALAMUD *et al.*, 1998). Two common types of binning are linear and logarithmic.

#### 2.3.1 Linear binning

Linear binning means that each bin has the same width, $\Delta r$. For linear binning, the non-cumulative (equation 1) and cumulative (equation 2) functions are related by

$$C = \frac{c}{(d-1)\Delta r} \tag{3}$$

and

$$D = d - 1. \tag{4}$$

For a derivation of this relationship, see BURROUGHS and TEBBENS (in press).

2.3.2 Logarithmic binning

Logarithmic binning means that bin width increases as the object size increases so that the ratio between successive bin widths is constant. The log of this ratio we call $a$. For logarithmically binned data, the non-cumulative (equation 1) and cumulative (equation 2) functions are related by

$$C = \frac{c}{1 - 10^{-ad}} \tag{5}$$

and

$$D = d \tag{6}$$

(BURROUGHS and TEBBENS, in press). The exponents of both the cumulative function $N(r)$ and the non-cumulative function $n(r)$ are the same, $d = D$. Note that this is different from linearly binned data where the exponents for $N(r)$ and $n(r)$ are related by $D = d - 1$.

## 3. Upper Truncation

Relationships between logarithmically binned non-cumulative functions and cumulative functions are shown in Figure 1. The non-cumulative function is upper-truncated at $n = 1$, corresponding to object size $r_1 = 200$. Line 1 of Figure 1 represents a non-cumulative function that follows a power law (equation 1). Line 2 of Figure 1 is the associated cumulative function. The cumulative function has a characteristic fall-off at the large object sizes due to the upper truncation of the non-cumulative function. Line 3 represents the underlying power law (equation 2). In the absence of upper-truncation of the non-cumulative function, the cumulative function would follow line 3 (Figure 1).

The relationships described above for functions can be applied to data sets. The cumulative distribution of an upper-truncated data set may be used to determine the underlying power law. A data set is upper-truncated when the observed largest object size is less than $r_{N1}$, where $r_{N1}$ is the object size for which the cumulative function equals one (i.e., $N(r_{N1}) = 1$). Since each point in a cumulative distribution includes all larger objects, upper truncation of the distribution decreases the cumulative number associated with each object size. The difference between the underlying

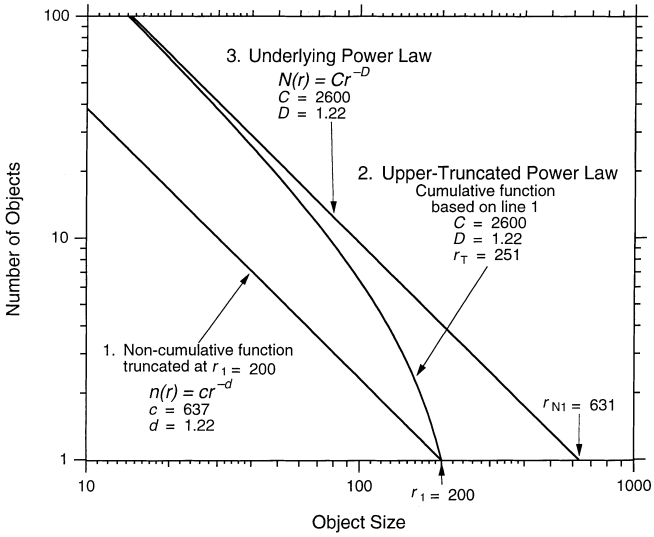


Figure 1

The relationship between log-binned non-cumulative and cumulative functions of an upper truncated power law. Line 1 is the non-cumulative function, $n(r) = cd^{-d}$ , where $c = 637$, $d = 1.22$, and $a = 0.1$ (see text, Section 2.3.2). This function is upper-truncated at $r_1 = 200$. Line 2 is the cumulative function associated with line 1 and has the form $M(r) = C(r^{-D} - r_T^{-D})$, where $C = 2600$, $D = 1.22$ and $r_T = 251$. The fall-off of line 2 at large object size is due to the non-cumulative function, line 1, being upper truncated. The term $r_T$ is the object size such that $M(r_T) = 0$. Line 3 represents the underlying power law. The underlying power law, $N(r) = Cr^{-D}$, can be determined by fitting $M(r)$ to line 2 to obtain the values of $C$ and $D$. If line 1 were not upper-truncated, Line 3 would be the cumulative function of line 1.

power law and the upper-truncated cumulative distribution we call the truncation parameter.

## 4. General Fitting Function

When a cumulative number-size distribution exhibits a fall-off from a power law at large object size due to upper-truncation, the proper function to fit to the distribution is an upper-truncated power law. The truncation parameter can be written as

$$N(r_T) = Cr_T^{-D}. \tag{7}$$

Subtracting the truncation parameter from the underlying power law yields the upper-truncated power law,

$$M(r) = C\left(r^{-D} - r_T^{-D}\right). \tag{8}$$

We call $M(r)$ the General Fitting Function. The term $r_T$ is a parameter determined by fitting equation (8) to the cumulative distribution and is the object size where $M(r)$

equals zero. We find the values of $C$, $D$, and $r_T$ that provide the best fit of equation (8) to the cumulative number-size distribution of a data set. The resulting values of $C$ and $D$ define the underlying power law. Since the number of objects in a cumulative distribution of a real data set must be an integer, and there will likely be scatter about a fitting function, $r_T$ may be greater than or less than the largest object size in the data set.

This approach of fitting the General Fitting Function to a cumulative number-size distribution to find the associated underlying power law we call the General Fitting Function (GFF) method.

## 5. *Analysis of Synthetic Discrete Data*

To demonstrate the GFF method, we created a synthetic data set that replicates characteristics similar to those observed for real discrete cumulative data. Starting with a cumulative power law, we calculated a set of values for $r$ such that for each successive $r$ the cumulative function decreases by one. We arbitrarily chose $C = 20$, $D = 1.4$, and truncated the data set at a maximum object size of $r_T = 3$ (Fig. 2). To create a synthetic cumulative data set, we binned the $r$ values linearly with $\Delta r = 0.1$. To fit a power-law function to a cumulative data set, previous authors (see Section 6) have fit a power function directly to a portion of the cumulative distribution. We call this approach the traditional method and present the results for the synthetic data in Figure 2A. The traditional method overestimates the value of $D$ in the cumulative power function. A more accurate approach is to use the GFF method. The best fit of $M(r)$ to the cumulative distribution is shown in Figure 2B. Figure 2C compares the

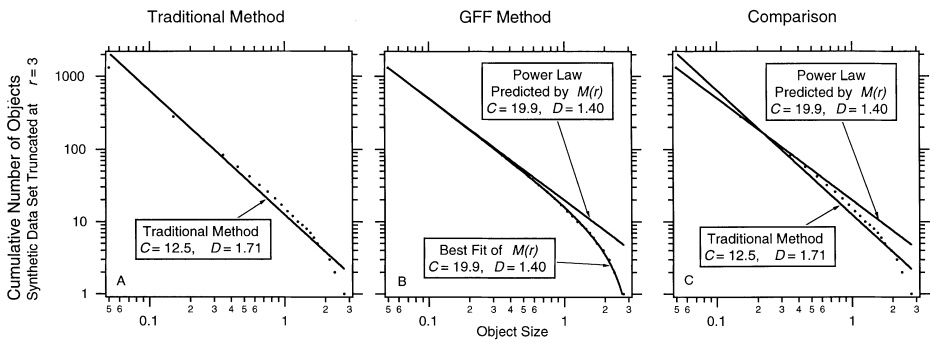


Figure 2

Synthetic cumulative data with $C = 20$ and $D = 1.4$ truncated at $r_T = 3$ for linear binning with $\Delta r = 0.1$. Cumulative data points are plotted at the beginning of each data bin. (A). Traditional method. A power-law fit to all data points provides incorrect values for $C$ and $D$. (B). The GFF method. The GFF fit to all data points produces $C = 19.9$ and $D = 1.4$, in good agreement with the underlying power law. (C). Comparison of traditional and GFF methods. The traditional method applied only to data points below the largest half-magnitude ($r \leq 1$ in this example) would provide better agreement between the two methods, although the traditional method would still overestimate $D$.

results of this fitting method to the results of the traditional method. The GFF method accurately determines $C$ and $D$ of the underlying power function.

## 6. Application of the GFF Method to Natural Systems

We apply the GFF method to four natural data sets. Previous authors have fit these data sets with a power function. The results of this section are summarized in Table 1.

### 6.1 Forest Fire Areas

The non-cumulative number-size distribution of forest fire areas in the Australian Capital Territory has been shown to follow a power law with $d = 1.49$ (MALAMUD *et al.*, 1998). We plot this forest fire data as a cumulative number-size distribution in Figure 3, and we fit a power law to this distribution by the traditional method (Fig. 3A). A best fit of $M(r)$ (Equation 8) to the same data is shown as the lower line in Figure 3B. The coefficients $C$ and $D$ found by the best fit of $M(r)$ to the cumulative distribution are used to plot the underlying power law, shown as a straight line in Figure 3B. The power-law functions found by fitting a power law to the cumulative distribution (Fig. 3A) and by fitting $M(r)$ to find the underlying power function (Fig. 3B) are compared in Figure 3C.

We make three significant observations from this comparison. First, the traditional method slightly overestimates $D$, ($D = 0.57$) compared to the GFF method ($D = 0.51$). Second, a comparison of $D$ determined by the GFF method ($D = 0.51$), to $d$ of the non-cumulative distribution ($d = 1.49$) (MALAMUD *et al.*, 1998), reveals that the difference between $D$ and $d$ is approximately 1, as expected for linearly binned data ($D = d - 1$). Third, since the data span several orders of magnitude, the discrepancy between the traditional and GFF methods is small. If the data were limited, for example by recording only forest fires with areas greater than 10 km$^2$, the discrepancy between the two methods would be greater. Fitting $M(r)$ to

Table 1

*Comparison between traditional method and GFF method*

| Data Set | Traditional Method | | GFF Method | | |
|---|---|---|---|---|---|
| | $C$ | $D$ | $C$ | $D$ | $r_T$ |
| ACT forest fire areas | 66.3 | 0.57 | 70.7 | 0.51 | 725 |
| Vernejoul coal field fault offset lengths | 176 | 0.559 | 205 | 0.468 | 386 |
| Frio strand plain exploration play oil and gas field sizes | 922 | 0.83 | 458 | 0.51 | 868 |
| Lengths of normal faults on the plains of Venus | $1.18 \times 10^8$ | 1.56 | $1.09 \times 10^6$ | 1.00 | $3.00 \times 10^4$ |

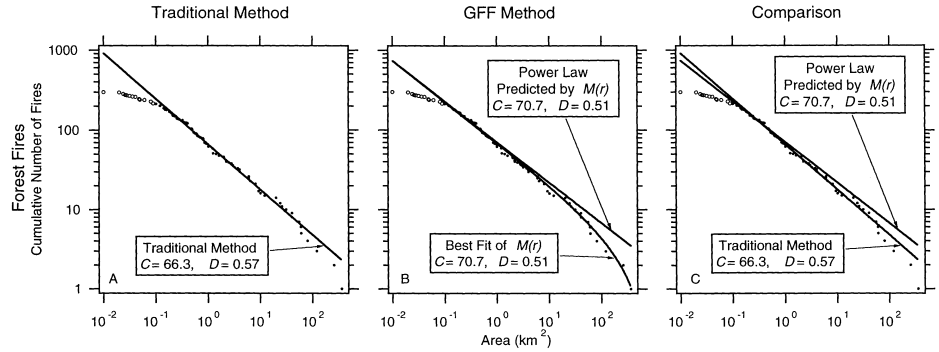


Figure 3

Cumulative distribution of forest fire areas for the Australian Capital Territory (1926–1991). (A). Traditional method. Closed circles (all values $> 0.10$ km$^2$) are fit with a power law yielding $D = 0.57$. (B). GFF method. Closed circles (all values $> 0.10$ km$^2$) are fit with $M(r)$ (curved line) to determine the underlying power law (straight line) yielding $D = 0.51$. This is consistent with the result of MALAMUD *et al.* (1998) who fit a power law to the non-cumulative distribution and found $d = 1.49$. (C). Comparison of traditional and GFF methods. The traditional method slightly overestimates $D$.

the cumulative distribution would determine the same coefficient and exponent of the underlying power law.

### *6.2 Fault Offset Lengths in the Vernejoul Coal Field*

Figure 4 shows the cumulative number-size distribution of fault offset lengths in the Vernejoul coal field as reported by VILLEMIN *et al.* (1995). VILLEMIN *et al.* fit a

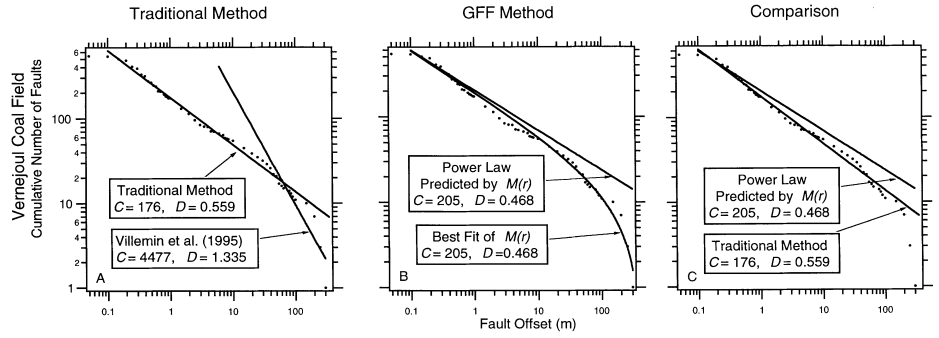


Figure 4

Cumulative distribution of fault offset lengths in the Vernejoul coal field from VILLEMIN *et al.* (1995). (A). Traditional method. Fitting a power law to all cumulative data points yields $D = 0.559$. Fitting of a power law to only the longest fault offsets yields $D = 1.335$ as reported by VILLEMIN *et al.* (1995). (B). GFF method. Data points are fit with $M(r)$ (curved line) to determine the underlying power law (straight line) yielding $D = 0.468$. (C). Comparison of traditional and GFF methods. The traditional method slightly overestimates $D$ whereas fitting a power law to the longest offsets seriously overestimates $D$.

power function to the longest offset lengths, where the record is likely to be most complete, and obtained $C = 4{,}477$ and $D = 1.335$ (Fig. 4A). For comparison purposes, we fit a power function to the cumulative distribution in the range $0.1 \leq r \leq 100$ resulting in $C = 176$ and $D = 0.559$ (Fig. 4A). Applying the GFF method, $M(r)$ is fit to the cumulative distribution yielding $C = 205$ and $D = 0.468$ for the underlying power law. Both the best fit of $M(r)$ and the underlying power law are shown in Figure 4B. The power functions found by the traditional method (Fig. 4A) and by the GFF method (Fig. 4B) are compared in Figure 4C.

We can draw three conclusions from this comparison (Fig. 4C). First, the approach of fitting the largest object sizes (Fig. 4A, $D = 1.335$) does not provide the best power-law fit to the data set. Second, while fitting a power function to the straightest portion of the cumulative distribution ($D = 0.559$) more closely approaches the underlying power law ($D = 0.468$) than fitting to only the largest objects ($D = 1.335$), the traditional method still overestimates $D$. Third, when fitting object sizes over three orders of magnitude, $M(r)$ and a power law are in closer agreement than is seen in many data sets that span fewer orders of magnitude. Since $M(r)$ accounts for the fall-off observed at large object size, we believe $M(r)$ more accurately characterizes the data.

### *6.3 Oil and Gas Accumulations in the Frio Strand Plain Exploration Play*

Figure 5 shows the cumulative number-size distribution of oil and gas field sizes discovered in the Frio Strand Plain exploration play, onshore Texas, through 1985. Figure 5A indicates the cumulative distribution and the power-law fit obtained by

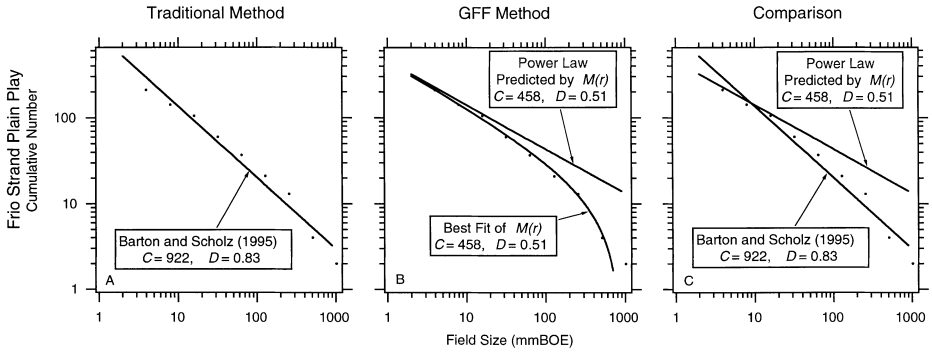


Figure 5

The cumulative distribution of oil and gas field sizes discovered in the Frio Strand Plain exploration play, onshore Texas, through 1985. (A). Traditional method. Fitting a power law to all plotted cumulative data points yields $D = 0.83$ as reported by BARTON and SCHOLZ (1995). (B). GFF method. The plotted data points used by Barton and Scholz are fit with $M(r)$ (curved line) to determine underlying power law (straight line) yielding $D = 0.51$. (C). Comparison of traditional and GFF methods. The traditional method overestimates $D$, which has a significant implication for the volume of undiscovered oil and gas (see text).

BARTON and SCHOLZ (1995) with $D = 0.83$. Figure 5B shows the fit of $M(r)$ to the same data, yielding $D = 0.51$ for the underlying power law. Figure 5C compares the results of BARTON and SCHOLZ (1995) and the GFF approach. As in the previous examples, the traditional approach overestimates the value of $D$ in the underlying power function. If the power function is used to predict the total volume of hydrocarbon accumulations (proportional to the area under the curve within any given volume range), the slope is very significant to the volume of assessed undiscovered resources. Even small differences in slope can produce large differences in the predicted total hydrocarbon volumes.

### *6.4 Fault Lengths on the Plains of Venus*

The non-cumulative number-size distribution for the lengths of normal faults on the plains of Venus has been shown to follow a power law (SCHOLZ, 1997). We plot this fault length data as a cumulative number-size distribution in Figure 6. A power law is fit to the straight portion of the graph in the middle of the data range by the traditional method, yielding the values $C = 1.18 \times 10^8$ and $D = 1.56$ (Fig. 6A). Applying the GFF method the best fit of $M(r)$ and its predicted underlying power law, with $C = 1.09 \times 10^6$ and $D = 1.00$, are shown in Figure 6B. The traditional and GFF methods are compared in Figure 6C.

SCHOLZ (1997) fit a power law directly to the non-cumulative distribution for this data set, with bin width of 1,667 m, and found $d = 2.02$ and $c = 1.82 \times 10^9$. We apply the GFF method to the cumulative distribution of the same data set. Using the values for $C$ and $D$ predicted by $M(r)$, and setting $\Delta r = 1{,}667$, equations (3) and (4)

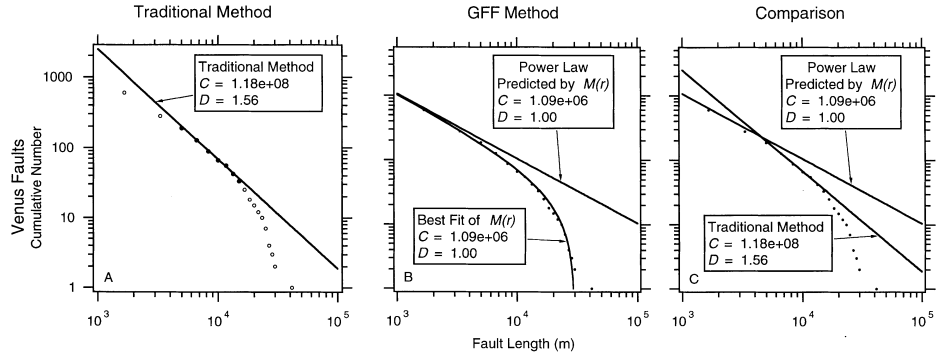


Figure 6

Cumulative distribution of fault lengths for normal faults on the plains of Venus. (A). Traditional method. Fitting a power law to the straight region in the middle of the data range (closed circles) yields $D = 1.56$. (B). GFF method. Data points are fit with $M(r)$ (curved line) to determine the underlying power law (straight line) yielding $D = 1.00$. (C). Comparison of traditional and GFF methods. The GFF method yields results consistent with the results of SCHOLZ (1997) for the associated non-cumulative distribution (see text).

predict the values $d = 2.00$ and $c = 1.82 \times 10^9$ for the non-cumulative distribution. The results obtained by applying the GFF method to the cumulative distribution are consistent with the results SCHOLZ (1997) obtained by fitting a power law to the non-cumulative distribution.

The Venus fault data are "about an order of magnitude better than any obtained on earth, both in range and the number of faults" (SCHOLZ, 1997). The high quality of the Venus data allowed SCHOLZ (1997) to fit a power law directly to the non-cumulative distribution. Generally, cumulative distributions are used to find power-law behavior of data sets. The Venus fault data set demonstrates two aspects of the GFF method. First, the traditional approach of fitting a power law directly to the cumulative distribution fails to find the correct scaling coefficient and exponent. The GFF method succeeds in obtaining results consistent with fitting the non-cumulative distribution (SCHOLZ, 1997). Second, for a cumulative distribution with a fall-off that is fit by $M(r)$, the GFF method accurately predicts the coefficients and exponents of the underlying power laws for both the cumulative and non-cumulative distributions.

## 7. Discussion

### 7.1 Previous Models

When seeking a power law to describe a cumulative number-size distribution, many authors have fit a power law directly to the distribution. For the many data sets where no fall-off is observed at large object size, this method is appropriate. Examples of data with a cumulative number-size distribution with little or no fall-off at large object size include cloud area in satellite and space shuttle imagery (PELLETIER, 1997), layer thickness of turbidites in the Kingston Peak formation (ROTHMAN *et al.*, 1994), and fault lengths observed near the East Pacific Rise at 13°–15°N (COWIE *et al.*, 1993). Lack of fall-off at large object size suggests that either the data set is not upper-truncated or that the largest observed object size is approximately equal to $r_{N1}$. The GFF method may be applied to data sets where no fall-off occurs. The result will be the same as fitting a power law directly to the distribution since when $r_T$ is large, $M(r)$ approaches the cumulative power law $N(r)$.

Several models account for the fall-off observed in cumulative number-size distributions at large object size (BARTON and SCHOLZ, 1995; LA BARBERA and ROTH, 1994; MALAMUD *et al.*, 1998; PICKERING *et al.*, 1995). In studying oil and gas accumulation sizes, BARTON and SCHOLZ (1995) found many distributions that exhibited a fall-off. The authors suggested that several power-law distributions of the same slope, offset from one another on a log–log plot, might be combined as a single distribution. This single distribution would exhibit a fall-off (BARTON and SCHOLZ,

1995, Fig. 2.6). We suggest that truncation of the data due to finite field size within a finite basin is a simpler explanation.

PICKERING *et al.* (1995, 1994) created synthetic cumulative number-size distributions and removed a few of the largest objects from the distribution (truncated or censored the data). The resulting cumulative distribution displayed a fall-off, with the amount of fall-off increasing as more of the largest objects were removed. To find the underlying power law in a natural cumulative distribution that exhibits fall-off, they estimated the number of objects missing above the upper end of the data set (their censoring correction) then applied that correction to the data. Fitting a power law to this corrected data gave a better approximation of $D$. The GFF method fits $M(r)$ directly to the cumulative distribution to yield the values for $C$ and $D$. The two methods should produce the same results because the term $Cr_T^{-D}$ that appears in $M(r)$ is equivalent to the censoring correction that PICKERING *et al.* (1995, 1994) approximated by iteration. We prefer the GFF method because it is simpler to apply.

In analyzing forest fire data, MALAMUD *et al.* (1998) took the derivative of the cumulative distribution, generally over five points, to create a non-cumulative distribution, and then fit a power law directly to the non-cumulative distribution. The results of the GFF method agree with the results obtained by Malamud *et al.* (1998). We again prefer the GFF method because it is not necessary to convert from a cumulative distribution to a non-cumulative distribution. Therefore, it is also not necessary to choose a bin width for the non-cumulative distribution.

A river network is a natural system that has been modeled as an upper-truncated power law. Strahler's ordering scheme and Horton's ratios have been used to derive equations describing river channel networks in a finite catchment area (LA BARBERA and ROTH, 1994 and references therein). These equations have been applied to river systems by other authors, for example CRAVE and DAVY (1997). The equations LA BARBERA and ROTH (1994) derived for analyzing river channel networks, have the same form as the General Fitting Function. The GFF method was developed independently by considering upper-truncated cumulative power-law distributions. The GFF method, not having been derived from laws describing a particular system, may be applied to a wide variety of natural systems (Section 6).

### *7.2 Truncation*

The upper truncation point, $r_T$, may be controlled by either the finite size of the study area, temporal limitation of the collected data, or constraints imposed by the underlying physics. Interpretation of $r_T$ depends on the individual study. Spatial truncation may result in the studies limited planar area (e.g., outcrop) or volume (e.g., petroleum basin). In the example of fault offset lengths, the largest measured offset cannot be larger than the dimensions of the sampled area. Thus, the finite

study area may produce truncation. Temporal truncation may occur when a data set spans a time interval that does not capture the largest naturally occurring event sizes. There may also be geological reasons, or controls from the underlying physics, that impose an upper limit to the possible observations in the study. For example, for continental rifts in Nevada, SCHOLZ and CONTRERAS (1998) found an upper limit to fault offset lengths of roughly 100 km controlled by resistive friction and the flexural restoring force. The authors point out that this truncation is not due to sampling limitations because all faults are well within their study region. In another example, maximum earthquake size has been found to be limited by the thickness of the rigid lithosphere (PACHECO *et al.*, 1992). Thus, in the cases of fault offset length and earthquake size, truncation may be explained by the underlying physics.

For the area of forest fires in the Australian Capital Territory we obtain a truncation value ($r_T$) of 725 km$^2$, which is roughly 30% of the total Territory area of 2,432 km$^2$. One interpretation of this result holds that the largest possible fires have not been observed in the record, consequently the record is truncated due to sampling. Alternatively, 725 km$^2$ may be a truncation that is controlled by the underlying physics and represents the maximum physically possible size of a forest fire in that region due to factors such as fuel supply and climate.

### *7.3 Plotting-binned Data*

In some cases, cumulative distributions are calculated from binned non-cumulative data sets (e.g., Venus faults in Section 6.4). Within the range of values in a bin, the value of $r$ chosen as the plotting location for the cumulative data affects the results. For cumulative distributions, each plotted point represents all values of the plotted size and larger. To correctly represent all objects within the data bin, the data points must be plotted at the value of $r$ corresponding to the beginning of the binning interval. This is illustrated in Figure 7 where the cumulative distribution for fault lengths on the plains of Venus is plotted both correctly at the beginning of each data bin and incorrectly at bin center. Fitting $M(r)$ to the correctly plotted data yields results consistent with those obtained by SCHOLZ (1997) for the non-cumulative distribution (see Section 6.4). When $M(r)$ is fit to the incorrectly plotted data, the underlying power law is incorrectly predicted to have values $C = 2.33 \times 10^7$ and $D = 1.35$.

For non-cumulative linearly binned data, plotting the number of objects in a bin at the bin's center is a good approximation as long as the width of the bin is small relative to the size of objects in the bin (BURROUGHS and TEBBENS, in press). For non-cumulative logarithmically binned data, the number of objects in a bin should be plotted at the start of the bin (Burroughs and Tebbens, in press). Failure to follow these plotting guidelines can lead to inaccurate fitting and misinterpretation of the data.

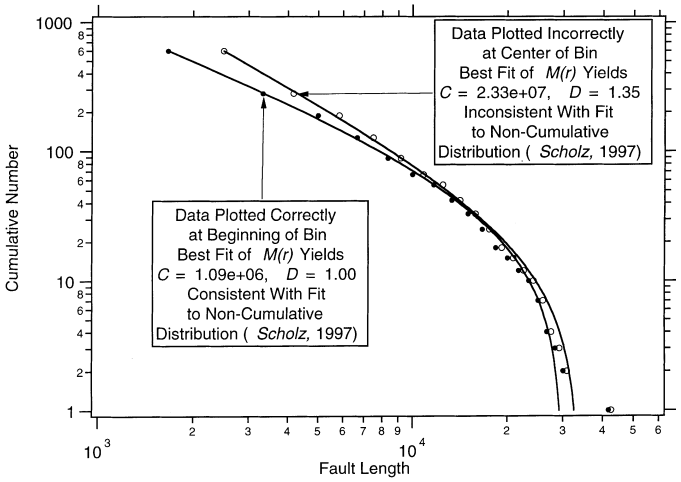


Figure 7

Proper method for plotting cumulative data that was calculated from binned non-cumulative data. Data points for the cumulative distribution of fault lengths on Venus are plotted correctly (as in Fig. 6) at the length corresponding to the beginning of each data bin (closed circles). The same data points are plotted incorrectly at the length corresponding to the center of each data bin (open circles). Fitting $M(r)$ to the incorrectly plotted data yields an erroneous prediction for the underlying power law.

### *7.4 Summary of GFF Method*

The GFF method finds the power function underlying upper-truncated cumulative number-size distributions. This method may be applied to cumulative distributions that exhibit a fall-off at large object size. A summary of the GFF method follows:

1. When attempting to find an underlying power function from a cumulative number-size distribution, the GFF method determines the coefficient and exponent of that function. When fall-off is observed in the cumulative distribution, the traditional method of fitting a power law directly to the cumulative number-size distribution estimates too negative an exponent and may over- or underestimate the coefficient for the underlying power function.
2. Fall-off of the cumulative number-size distribution occurs when data are truncated at large object sizes, as shown in this and earlier work (e.g., BROOKS and ALLMENDINGER, 1996; PICKERING *et al.*, 1995). The GFF method provides a function that fits this fall-off. In addition to providing the coefficient and exponent of the underlying power function, the GFF method characterizes the upper truncation.
3. Different methods of binning the same data (e.g., linear or logarithmic) produce different non-cumulative distributions, whereas the cumulative distributions will be the same. Therefore, the interpretation is less complicated if the

cumulative number-size distribution is used to determine the underlying power function.

## 8. Conclusions

In this work the GFF method was developed to fit upper-truncated cumulative power law distributions. The GFF is shown to be applicable to forest fire area, fault offset lengths, hydrocarbon accumulations, and fault lengths. Thus, these four systems have characteristics that may be described by upper-truncated power laws. Data sets describing other natural systems may also exhibit fall-off at large object size. The results of this work suggest that many natural systems may be characterized by upper-truncated power laws.

## 9. Notation

| Symbol | Definition |
|---|---|
| $a$ | log of ratio between successive bin widths for logarithmically binned data |
| $c$ | coefficient of non-cumulative power law |
| $C$ | coefficient of cumulative power law |
| $-d$ | exponent of non-cumulative power law |
| $-D$ | exponent of cumulative power law |
| $M(r)$ | General Fitting Function (GFF): continuous function describing any power-law based upper-truncated cumulative number-size distribution |
| $n(r)$ | continuous function describing a non-cumulative number-size distribution of data |
| $N(r)$ | continuous function describing a cumulative number-size distribution of data |
| $r$ | object size |
| $\Delta r$ | sampling interval or bin width |
| $r_{\mathrm{T}}$ | object size for which the upper-truncated power law equals zero |
| $r_{\mathrm{N1}}$ | object size where the cumulative power law equals 1 |

## Acknowledgements

Kevin Dove and Emilio Toro of the University of Tampa provided invaluable assistance reviewing the mathematical analysis. This manuscript benefited from discussions with Bruce Malamud, Chris Scholz, Colin Stark, and Donald Turcotte, and the suggestions of one anonymous reviewer. Chris Barton provided the Frio Strand Plain data, and his many constructive reviews and helpful suggestions greatly

improved the manuscript. Bruce Malamud and Chris Scholz generously provided the forest fire data and the Venus fault data, respectively.

## References


Barton, C.C., and Scholz, C.H., *The fractal size and spatial distribution of hydrocarbon accumulations*. In *Fractals in Petroleum Geology and Earth Processes* (eds. C.C. Barton and P.R. LaPointe) (Plenum, New York 1995).

Brooks, B.A., and Allmendinger, R.W. (1996), *Fault spacing in the El Teniente Mine, Central Chile: Evidence for Nonfractal Geometry*, J. Geophys. Res. *101*, 13,633–13,653.

Burroughs, S.M., and Tebbens, S.F. (in press), *Upper-truncated Power Law Distributions*, Fractals.

Cowie, P.A., Scholz, C.H., Edwards, M., and Malinverno, A. (1993), *Fault Strain and Seismic Coupling on Mid-ocean Ridges*, J. Geophys. Res. *98*, 17,911–17,920.

Crave, A., and Davy, P. (1997), *Scaling Relationships of Channel Networks at Large Scales: Examples from two Large-magnitude Watersheds in Brittany, France*, Tectonophysics *269*, 91–111.

Feder, J., *Fractals* (Plenum Press, New York 1988).

Gutenberg, B., and Richter, C.F., *Seismicity of the Earth and Associated Phenomena* (Princeton University Press, Princeton 1949).

Jackson, P., and Sanderson, D.J. (1992), *Scaling of Fault Displacements from the Badajoz-Cordoba Shear Zone, SW Spain*, Tectonophysics *210*, 179–190.

La Barbera, P., and Roth, G. (1994), *Invariance and Scaling Properties in the Distributions of Contributing Area and Energy in Drainage Basins*, Hydrological Proceses *8*, 125–135.

Lu, C., Vere-Jones, D., and Takayasu, H. (1999), *Avalanche Behavior and Statistical Properties in a Microcrack Coalescence Process*, Phys. Rev. Lett. *82*, 347–350.

Malamud, B.D., Morein, G., and Turcotte, D.L. (1998), *Forest Fires: An Example of Self-organized Critical Behavior*, Science *281*, 1840–1842.

Malamud, B.D., Turcotte, D.L., and Barton, C.C. (1996), *The 1993 Mississippi River Flood: A One Hundred or a One Thousand Year Event*?, Environm. and Engin. Geoscience *2*, 479–486.

Malinverno, A. (1997), *On the Power-law Size Distribution of Turbidite Beds*, Basin Res. *9*, 263–274.

Mandelbrot, B.B., *Fractals and Scaling in Finance* (Springer, New York 1997).

Maritan, A., Rinaldo, A., and Rigon, R. (1996), *Scaling Laws for River Networks*, Phys. Rev. E *53*, 1510–1515.

Nagatani, T. (1993), *Crossover Scaling in Scheidegger's River-network Model*, Phys. Rev. E *47*, 3896–3899.

Pacheco, J., Scholz, C.H., and Sykes, L.R. (1992), *Changes in Frequency-size Relationship from Small to Large Earthquakes*, Nature *355*, 71–73.

Pelletier, J. (1997), *Kardar-Parisi-Zhang Scaling of the Height of the Convective Boundary Layer and Fractal Structure of Cumulus Cloud Fields*, Phys. Rev. Lett. *78*, 2672–2675.

Pelletier, J.D., Malamud, B.D., Blodgett, T., and Turcotte, D.L. (1997), *Scale- invariance of Soil Moisture Variability and its Implications for the Frequency-size Distribution of Landslides*, Engin. Geology *48*, 255–268.

Pickering, G., Bull, J.M., and Sanderson, D.J. (1995), *Sampling Power-law Distributions*, Tectonophysics *248*, 1–20.

Pickering, G., Bull, J.M., Sanderson, D.J., and Harrison, P.V., *Fractal fault displacements: A case study from the Moray Firth, Scotland*. In *Fractals and Dynamic Systems in Geoscience* (ed. J.H. Kruhl) (Springer-Verlag, Berlin 1994) pp. 105–118.

Rothman, D.H., Grotzinger, J.P., and Flemings, P. (1994), *Scaling in Turbidite Deposition*, J. Sedim. Res. *A64*, 59–67.

Scholz, C.H. (1997), *Earthquake and Fault Populations and the Calculation of Brittle Strain*, Geowissenschaften *15*, 124–130.

Scholz, C.H. (1998), *A Further Note on Earthquake Size Distributions*, Bull. Seismol. Soc. Am. *88*, 1325–1326.

SCHOLZ, C.H., and CONTRERAS, J.C. (1998), *Mechanics of Continental Rift Architecture*, Geology *26*, 967–970.

SORNETTE, A., DAVY, P., and SORNETTE, D. (1993), *Fault Growth in Brittle-ductile Experiments and the Mechanics of Continental Collisions*, J. Geophys. Res. *98*, 12,111–12,139.

TURCOTTE, D.L., *Fractals and Chaos in Geology and Geophysics* (Cambridge Press, Cambridge 1997).

VILLEMIN, T., ANGELIER, J., and SUNWOO, C., *Fractal distribution of fault length and offsets: Implications of brittle deformation evaluation – The Lorraine coal basin*. In *Fractals in the Earth Sciences* (eds. C. Barton and P.R. LaPointe) (Plenum Press, New York 1995) pp. 205–226.

WANG, J.-H. (1995), *Effect of Seismic Coupling on the Scaling of Seismicity*, Geophys. J. Int. *121*, 475–488.

(Received December 10, 1999, accepted May 4, 2000)