A fractal-based approach to lake size-distributions

David A. Seekell,1 Michael L. Pace,1 Lars J. Tranvik,2 and Charles Verpoorter3

Received 26 October 2012; revised 1 January 2013; accepted 3 January 2013; published 20 February 2013.

[1] The abundance and size distribution of lakes is critical to assessing the role of lakes in regional and global biogeochemical processes. Lakes are fractal but do not always conform to the power law size-distribution typically associated with fractal geographical features. Here, we evaluate the fractal geometry of lakes with the goal of explaining apparently inconsistent observations of power law and non–power law lake size-distributions. The power law size-distribution is a special case for lakes near the mean elevation. Lakes in flat regions are power law distributed, while lakes in mountainous regions deviate from power law distributions. Empirical analyses of lake size data sets from the Adirondack Mountains in New York and the flat island of Gotland in Sweden support this finding. Our approach provides a unifying framework for lake size-distributions, indicates that small lakes cannot dominate total lake surface area, and underscores the importance of regional hypsometry in influencing lake size-distributions. Citation: Seekell, D. A., M. L. Pace, L. J. Tranvik, and C. Verpoorter (2013), A fractal-based approach to lake size-distributions, Geophys. Res. Lett., 40, 517–521, doi:10.1002/grl.50139.

1. Introduction

[2] How many lakes are there and how big are they? This is one of the most fundamental questions when assessing the roles of lakes in regional and global biogeochemical cycling. Small lakes are generally not recorded on maps, and even the best compilations of global lake data are thought to greatly underestimate the abundance and surface area of small lakes [Meybeck, 1995; Lehner and Döll, 2004; Downing et al., 2006]. Consequently, the abundance and surface area of small, unrecorded lakes is typically estimated based on extrapolations from power law size-distributions [Downing et al., 2006]. Analyses based on this methodology have revealed that lakes cover a much greater portion of Earth’s land surface (~3%) than previously believed, that lakes store substantial amounts of carbon in their sediments (up to 820 Pg C), and that greenhouse gas emissions from lakes may almost completely offset the terrestrial carbon sink [e.g., Cole et al., 2007; Tranvik et al., 2009; Bastviken et al., 2011]. Because of their abundance and high biogeochemical rates, small lakes appear to play a large role in carbon emission and sequestration [Wetzel, 1990; Downing, 2010].

[3] There are two principal lines of evidence in support of extrapolation based on a power law lake size-distribution. First, lakes are fractals, meaning the convolutedness of their shorelines is proportional to the scale at which they are examined [Goodchild, 1988; Hamilton et al., 1992]. Fractal geological features, like lakes, typically conform to a power law size-distribution [Mandelbrot, 1983]. Second, linear regressions on log-abundance log-size plots typically have high r^2 values, a pattern consistent with power law-distributed data [Downing et al., 2006; Seekell and Pace, 2011]. However, many size distributions have high r^2 values on log-abundance log-size plots when small values are excluded (i.e., lower size limits are truncated because of the uncertain accuracy of observations at these lake sizes). Some high-resolution lake size data sets that accurately observe small lakes have low r^2 values and deviate considerably from the power law distribution, potentially indicating orders of magnitude overestimations of the abundance of small lakes by power law distributions [Meybeck, 1995; Seekell and Pace, 2011]. For instance, a recent lake census for the United States has found that the power law distribution does not adequately describe the size distribution of lakes in some regions and that early extrapolations based on the power law distribution may have overestimated the global abundance of lakes by 240 million [McDonald et al., 2012]. These differences in estimates of lake abundance are significant. For example, Lewis [2011] compared estimates of global gross primary production of lakes based on power law and non–power law lake size-distributions. The power law–based estimate produced a 45% larger estimate than an alternate non–power law distribution. Hence, the size distribution of lakes is poorly constrained, but understanding lake size-distributions is critical to evaluating the role of lakes in regional and global biogeochemical cycles [Tranvik et al., 2009].

[4] Analyses of lake size-distributions [e.g., Hamilton et al., 1992; Downing et al., 2006; Seekell and Pace, 2011; McDonald et al., 2012] are limited, and there is a critical lack of theory from which to derive testable hypotheses to guide new developments in global-scale limnological analyses. Here, we consider lake size-distributions in a fractal geometry framework, with the goal of resolving inconsistent observations of power law and non–power law lake size-distributions. We specifically focus on regional hypsometry (area–elevation relationships) in influencing the shape of lake size-distributions. We evaluate our findings with analyses of lake size-distributions from mountainous and flat regions.

2. Conceptual Foundation for Lake Size-Distributions in Fractal Geometry

[5] A lake shoreline is equivalent to a contour line for the lake surface elevation. Consequently, fractal geometry theory
developed for topographic contour lines is applicable to the analysis of lake shorelines. Here, we synthesize evidence from several studies of the fractal geometry of contour lines that are relevant to lakes and show that apparently conflicting reports of both power law and non–power law lake size-distributions are consistent with expectations from fractal geometry [Downing et al., 2006; Seekell and Pace, 2011].

If we approximate a landscape with a fractal surface [Mandelbrot, 1975, 1983; Russ, 1994] and intersect the landscape with a horizontal plane at the mean landscape elevation, the points where fractal surface returns to the horizontal plane form a fractal known as a random Cantor set [Matsushita et al., 1991; Russ, 1994]. The fractional dimension of the random Cantor set is one less than the fractal dimension of the surface and can be measured based on the distribution of distances between returns [Russ, 1994]. The return points can be connected to form contour lines. Longer distances between returns lead to larger areas enclosed by the contour lines [in the sense of Matsushita et al., 1991; Russ, 1994]. Because lake shorelines are contour lines for the lake surface elevation, the distance between returns is analogous to lake areas [in the sense of Matsushita et al., 1991; Russ, 1994]. The shorelines of individual lakes have a fractal dimension, but the collections of areas enclosed by the shorelines are also power law distributed and associated with a fractal dimension that represents the collection of shorelines [Matsushita et al., 1991; Isogami and Matsushita, 1992; Russ, 1994; Sasaki et al., 2006]. The fractional dimension of the size distribution of lakes near the mean elevation is measured with the regression

\[ N = c A^{-b} \]  \hspace{1cm} (1)

where \( N \) is the number of lakes greater than or equal to the area \( A \), \( c \) is a constant, \( b = D/2 \), and the functional form (i.e., power law form) of the regression is based on the first return rate of a fractional Brownian to the mean elevation [Goodchild, 1988; Matsushita et al., 1991]. \( D \) is the fractal dimension of the shorelines surrounding the lake area and is constrained between \( D = 1 \) (a population of perfectly smooth shorelines) and \( D = 2 \) (a population of shorelines so irregular they are space filling). Hence, there is a theoretical basis for a power law size-distribution of lakes at the mean elevation, but there are theoretical constraints \((0.5 \leq b \leq 1)\) on the plausible range of exponents [Goodchild, 1988; Hamilton et al., 1992].

In some landscapes (e.g., mountainous ones) lakes are present at elevations far from the mean [Goodchild, 1988]. If returns through a horizontal plane intersecting the landscape at an elevation far from the mean are recorded, the distribution of return times through the section (lake areas) begins to deviate from a power law [cf. Ding and Yang, 1995]. In this case the log-abundance log-size equation takes the form

\[ N = c A^{-b} \times \exp(-d \times A) \]  \hspace{1cm} (2)

where \( d \) is a constant and the functional form of the regression is determined by the probability of the first return of a fractional Brownian motion to an elevation not equal to the mean [Ding and Yang, 1995]. This second size distribution is just a more generalized function than the equation for the mean elevation only. If \( d = 0 \), the equation is equivalent to the equation for the size distribution at the mean elevation (i.e., equation (1)). Conceptually, this means that in regions of high vertical relief, there is less surface area at any one elevation, and hence, there is simply not enough surface area for lakes to form in a great enough abundance to achieve a power law. As a consequence, regional hypsometry likely plays a strong role in determining the shape of the size distribution of lakes.

3. Empirical Analysis

Based on the concepts outlined above, we conclude that while the shorelines of individual lakes are fractal, this only leads to a power law size-distribution of lakes in specific, although not necessarily uncommon, cases. To test these expectations, we analyzed the lake size data from the Adirondack Mountains in New York [Seekell and Pace, 2011] and from the island of Gotland in Sweden [Verpoorter et al., 2012]. We selected the Adirondack data because they derive from a well-defined mountainous region and the Gotland data because they derive from a well-defined flat region.

We evaluated the Adirondack data based on four criteria derived from the fractal concept for lakes. First, we tested the data for deviation, based on the \( r^2 \) value, from a power law on a log-abundance log-size plot [Seekell and Pace, 2011]. Power law distribution data should form a straight line on a log-abundance log-size plot, and a low \( r^2 \) value indicates deviation from a straight line and hence deviation from a power law [Seekell and Pace, 2011; Appendix A]. There is a large vertical relief in the Adirondack data set, and hence, the distribution should deviate significantly from a power law. Second, we extracted lakes at the mean lake elevation \((n = 19 \text{ at elevation } = 503.6 \text{ m})\), which for this data set is approximately the same as the mean landscape elevation, and tested them, based on the \( r^2 \) value, on a log-abundance log-size plot for deviation from a power law. Lakes at the mean elevation should not deviate significantly from a power law distribution. Third, we compared the fractal dimension \( D \) from the log-abundance log-size plots of lakes at the mean elevation to fractal dimensions derived from dimensional analysis (log perimeter-log area analysis) [Russ, 1994]. The values estimated from the log-abundance log-size plots should be theoretically plausible and similar to estimates of \( D \) from dimensional analysis. Fourth, we compared the fit of size-distribution equations (1) and (2) (above) to lakes away from the mean elevation. For lakes away from the mean elevation, size-distribution equation (2) should exhibit improved fit relative to size-distribution equation (1).

The Adirondack Mountain data set is based on a stratified sample of lakes digitized from USGS topographical maps designed to accurately represent lakes of different sizes and includes ponds as small as 0.1 ha. For the Adirondack Mountain lake data set \((n = 1469; \text{Figure 1})\), the slope of the log-abundance log-size regression on the entire data set was \(-0.658 \text{ (Figure 1a)}\). This falls within the theoretical constraints of \(0.5 \leq b \leq 1\), but the \( r^2 \) value \((r^2 = 0.853)\) was significantly lower than expected (critical \( r^2 = 0.990)\) if the data conformed to a power law distribution [Seekell and Pace, 2011]. The lake elevation distribution roughly conformed to the normal distribution (Figure 1b), and there was considerable variability in lake elevation (mean elevation = 503.6 m, standard deviation = 111.7). The slope of the log-abundance log-size regression for lakes at the mean elevation (503 m, \(n = 19 \text{ lakes})\) was \(-0.613 \text{ (Figure 1c)}\). This slope falls within
the theoretical constraints of $0.5 \leq b \leq 1$, and the $r^2$ value ($r^2 = 0.846$) was consistent with data that conform to a power law distribution [Seekell and Pace, 2011]. The fractal dimension of the size distribution is $D = 1.23$, which is very similar to the fractal dimension ($D = 1.22$) derived from dimensional analysis for the entire lake data set.

[13] We fit the size-distribution equations (1) and (2) to lake sizes from a 25 m range ($n = 67$ lakes between 612.5 and 637.5 m) about 100 m above the mean elevation. Using lakes from this small range as opposed to just from one elevation was necessary in order to achieve a sample size large enough for analysis, and the range of elevations was based on the bounds of an arbitrarily selected bin from a histogram of lake elevations. We compared the fits of the alternate regression models on log-abundance log-size plots by examining the dual criteria of linearity of predicted versus observed values and evenness of distribution of points above and below the regression line [Quandt, 1964]. Both equations are unbiased in the statistical sense (the slopes of regressions on these variables = 1), but the predicted and observed values are not linearly related for size-distribution equation (1) (Figure 2a) whereas they are for size-distribution equation (2) (Figure 2b). The points are much more evenly distributed around the regression line for size-distribution equation (2) (Figure 2b) than they are for size-distribution equation (1) (Figure 2a). Based on these dual criteria, the fit of size-distribution equation (1) ($r^2 = 0.826$) was poor relative to the fit of size-distribution equation (2) ($r^2 = 0.992$). We do not attempt to interpret the values of the coefficients from size-distribution equation (2) because both independent variables are lake surface area and this colinearity can lead to highly inaccurate parameter estimates.

[14] The Gotland lake ($n = 114$) data are based on a recent high-resolution census of lakes greater than 0.01 km², described in detail by Verpoorter et al. [2012]. We tested the lake area for deviation from a power law on a log-abundance log-size plot based on the $r^2$ value. Gotland is an island with low vertical relief, and hence, the data should not deviate from the power law. We compared the fractal dimension derived from the log-abundance log-size regression to a fractal dimension derived from dimensional analysis. The fractal dimensions should be similar to each other. For the Gotland data set, the slope of the log-abundance log-size regression on the entire data set was $-0.795$ (Figure 3a). This falls within the theoretical constraints of $0.5 \leq b \leq 1$, and the $r^2$
ancient lake basins on Mars (Appendix B).

Fractal concepts apply to lakes, is supported by additional empirical analyses of size-distribution. The fractal concept is not specific to Earth’s surface, and the generality of the concept, as applied to lakes, is supported by additional empirical analyses of ancient lake basins on Mars (Appendix B).

Figure 3. (a) Log-abundance log-size plot for lakes in Gotland. The slope of the regression line is \(-0.795\). The distribution does not depart significantly from the power law distribution. (b) Histogram of lake elevations in Gotland. All lakes are near the mean elevation (mean elevation = 18.8 m; standard deviation = 18.7).

The log-abundance log-size plot was consistent with that expectation from independent measurements of the topography. However, some distributions, even for lakes at or near the mean elevation, might still deviate from the power law distribution. In many distributions the upper tail deviates from a power law because it is impossible to fit enough large lakes on a finite surface, truncating the distribution. In this case the lower tail of the distribution will still potentially conform to the power law distribution [Hamilton et al., 1992]. When small lakes are not completely enumerated, the lower tail of the size distribution may depart from a power law, but only below the minimally reliably mapped area. Deviations in the lower tail above the minimum reliably mapped size are not the results of mapping error [Seekell and Pace, 2011]. We cannot completely rule out potential impacts of truncation or mapping error on our analysis. However, we observed curvature extending throughout the entire Adirondack lake distribution (e.g., Figure 1a). This curvature is inconsistent with a power law, but consistent with size-distribution equation (2) [Seekell and Pace, 2011]. Size-distribution equation (2) cannot explain the complete flattening of the extreme lower tail (lakes < 0.01 km\(^2\)), which could be due to omission of small lakes from the data set. This flattening could also occur if scale-dependent geomorphic processes have eliminated these very small lakes from the landscape. Our approach serves as null hypotheses for landscapes without these processes and hence does not account for this effect [Goodchild, 1988]. We did not observe any patterns (e.g., breaks in linearity, regular spacing in between points on the low end of area and perimeter ranges) in our perimeter-area relationships that would suggest an adverse effect of mapping resolution.

4. Discussion

Our empirical analysis of lake sizes from a flat and mountainous region supported expectations drawn from the fractal concept and can explain apparently conflicting observations of power law and non-power law lake size-distributions. Regional hypsometry influences the shape of lake size-distributions such that mountainous regions likely depart from the power law lake size-distribution, whereas other flatter regions likely conform to a power law lake size-distribution. The fractal concept is not specific to Earth’s surface, and the generality of the concept, as applied to lakes, is supported by additional empirical analyses of ancient lake basins on Mars (Appendix B).

Our empirical results for lakes at the mean elevation met the dual criteria for the fractal concept that (1) the shape on a log-abundance log-size plot conformed to a power law distribution and (2) the slope of the regression on the log-abundance log-size plot was consistent with that expectation from independent measurements of the topography. However, some distributions, even for lakes at or near the mean elevation, might still deviate from the power law distribution. In many distributions the upper tail deviates from a power law because it is impossible to fit enough large lakes on a finite surface, truncating the distribution. In this case the lower tail of the distribution will still potentially conform to the power law distribution [Hamilton et al., 1992]. When small lakes are not completely enumerated, the lower tail of the size distribution may depart from a power law, but only below the minimally reliably mapped area. Deviations in the lower tail above the minimum reliably mapped size are not the results of mapping error [Seekell and Pace, 2011]. We cannot completely rule out potential impacts of truncation or mapping error on our analysis. However, we observed curvature extending throughout the entire Adirondack lake distribution (e.g., Figure 1a). This curvature is inconsistent with a power law, but consistent with size-distribution equation (2) [Seekell and Pace, 2011]. Size-distribution equation (2) cannot explain the complete flattening of the extreme lower tail (lakes < 0.01 km\(^2\)), which could be due to omission of small lakes from the data set. This flattening could also occur if scale-dependent geomorphic processes have eliminated these very small lakes from the landscape. Our approach serves as null hypotheses for landscapes without these processes and hence does not account for this effect [Goodchild, 1988]. We did not observe any patterns (e.g., breaks in linearity, regular spacing in between points on the low end of area and perimeter ranges) in our perimeter-area relationships that would suggest an adverse effect of mapping resolution.

Our analysis is based on a theoretical fractal surface. Fractal surfaces can approximate a wide variety of landscapes [Mandelbrot, 1975, 1983], but are scale free and hence do not include scale-dependent geological processes including some that may change lake abundances [Goodchild, 1988]. The fractal approach utilized here is only an approximation to real landscapes, but this approach is advantageous for simplifying the development of testable hypotheses that are useful for regional and global limnological studies [Goodchild, 1988]. The geology of landscapes may modify how well data fit the simple fractal landscape model. For example, Gotland formed through isostatic uplift and the draining of Baltic Sea’s freshwater predecessors. The rock is carbonate, resulting in karst weathering, a landscape likely to mimic the equally peaked and pitted fractal surface given by Goodchild’s model. For example, Gotland formed through isostatic uplift and the draining of Baltic Sea’s freshwater predecessors. The rock is carbonate, resulting in karst weathering, a landscape likely to mimic the equally peaked and pitted fractal surface given by Goodchild’s model.
regions in our empirical analysis, it may not be appropriate for large-scale analyses that combine multiple physiographic regions. Overall, our empirical results were largely consistent with fractal expectations, suggesting that this approach will have utility for further analyses of lake morphometry and lake size-distributions.

In an analysis by Downing et al. [2006], the value of $b$ for the world’s largest lakes was 1.06. This has important implications because, if the data conform to a Pareto distribution (power law distribution), this slope ($b = 1.06$) indicates that small lakes dominate the total surface area covered by lakes. However, this value is inconsistent with fractal geometry theory in that the slope $b$ is constrained between $b = 0.5$ and $b = 1$ [Goodchild, 1988]. A potential reason for finding $b > 1$ may be truncation of the lower tail of the lake size-distribution. In their analysis of the world’s largest lakes, Downing et al. [2006] excluded all lakes less than 10 km² in area. This type of truncation can make data from many size distributions mimic the linearity of a power law distribution adequate for regional lake-rich landscapes, many size distributions mimic the linearity of a power law distribution in terms of regional lake dominance and pace [2005], but extrapolation from many of these estimates remains difficult because the mean and variance calculated from samples of power law-distributed data vary wildly with small changes in sample size [Mandelbrot, 1963]. Hence, there is tremendous uncertainty in (1) whether the power law distribution adequately describes the global and regional size distribution of lakes and (2) what the relative contribution of small versus large lakes is to global lake surface area. These uncertainties are probably best resolved by complete enumeration of lakes [Seekell and Pace, 2011; McDonald et al., 2012].

Most of Earth’s land surface (~75%) is located outside of mountainous or high-elevation regions, and consequently, the power law distribution may hold over much of Earth’s surface [Miller and Spoolman, 2011]. Many lake-rich regions are relatively flat (e.g., Finland), and power law distribution fits to lakes in these regions may be useful for understanding lake hydrological and biogeochemical response to environmental change. Small lakes are thought to play an important role in regional- and global-scale biogeochemical cycles, and the potential dominance of small lakes in terms of surface area has been cited as important to this argument [Downing, 2010]. While our results suggest that small lake dominance of surface area is unlikely, even in regions where the power law distribution holds (because $b$ must be $\leq 1$), this does not preclude small lakes from significance in regional biogeochemical cycles. Small lakes typically have higher fluxes and faster reaction rates than large lakes and consequently may still contribute disproportionately to biogeochemical cycles of lake-rich regions [Downing, 2010]. This potential importance, however, is not due to small lake dominance of total lake surface area. Many critical biogeochemical processes in lakes may be driven by processes at a regional scale [Lapiere and del Giorgio, 2012]. Relating the spatial organization of biogeochemical processes to regional lake size-distributions may be a promising route for improved understanding of the role of lakes in biogeochemical cycles at broad spatial scales.

References


Supplemental Material

Appendix A: Description of the $r^2$ test

Overview of the $r^2$ test

Power-law distributed data form a straight line on a log-abundance log-size plot [Gaudoin et al., 2003; Seekell and Pace, 2011]. The $r^2$ value from a regression on the log-abundance log-size plot is an indicator of linearity [Gaudoin et al., 2003; Seekell and Pace, 2011]. An $r^2$ value of 1 indicates a straight line while low $r^2$ values indicate a departure from linearity and hence a departure from the power-law distribution [Gaudoin et al., 2003; Seekell and Pace, 2011]. Interpreting $r^2$ values as deviation from a linearity hypothesis (null hypothesis of $r^2 = 1$, meaning a perfect linear relationship between log-abundance and log-size), as opposed to interpreting $r^2$ values as variance explained by the regression (i.e. a null hypothesis of $r^2 = 0$) allows for broad evaluations of log-abundance log-size regressions in the literature if the original data is no longer available because other statistical tests for power-law distribution are rarely applied to lake-size distributions, but the $r^2$ value and sample size for regressions on log-abundance log-size plots typically are reported.

This general approach - examining deviation from a straight line as a test for goodness-of-fit to a statistical distribution - is relatively common. For example, the probability plot correlation coefficient test is commonly applied to test for the normal distribution [Filliben, 1975]. This approach calculates a linear correlation coefficient between data and corresponding normal order statistics (i.e. calculates a correlation coefficient for data on a common probability plot). There will be a linear relationship ($r = 1$) if the data conform to a normal distribution, but the relationship will deviate from a straight $r < 1$ if the data deviate from a normal distribution [Filliben, 1975].
Significance testing with the $r^2$ test

The statistical significance of the $r^2$ test is evaluated by comparing the $r^2$ value from a regression on a log-abundance log-size plot to a table of empirical percentage points [Seekell and Pace, 2011]. The percentage points are calculated by simulating power-law distributed data of a certain sample size and calculating $r^2$ values on a log-abundance log-size plot. This process is replicated a large number of times in order to develop a distribution of $r^2$ values when the data conform to the power-law distribution. This is repeated for a wide variety of sample sizes. The observed $r^2$ value from an analysis is compared to the distribution of $r^2$ values from simulated data (where the sample size was equal to that of the empirical analysis). If the observed $r^2$ value is less than or equal to the 5th percentile of the simulated distribution of $r^2$, the data are considered significantly different from a power-law distribution because an $r^2$ value that low is unlikely if the data are from a power-law distribution. Seekell and Pace [2011] give a table of percentage points for evaluating the significance of $r^2$ values from log-abundance log-size plots for a variety sample sizes between 5 and 5000. We generated new critical values for significance for the specific sample sizes used in the present analysis. We describe this procedure in detail below.

Power analysis for the $r^2$ test

It is useful to know the statistical power of the $r^2$ to reject the null hypothesis of a power-law distribution. We simulated critical values for the $r^2$ test by generating 50,000 power-law distributed datasets of sample sizes $n = 19, 29, 114, \text{ and } 210$ and estimated the 5th percentile of $r^2$ values calculated from log-abundance log-size regressions on the simulated data [Seekell and Pace, 2011].
For \( n = 1469 \), 10,000 datasets were used. The sample sizes correspond to the sample sizes that the \( r^2 \) test is applied to in our empirical analysis. The simulation was replicated five times and the average of the five 5\(^{th} \) percentile values was taken as the critical value for the power analysis [Seekell and Pace, 2011]. We use these critical values (Table S1) to assess the significance of \( r^2 \) values in the main text as well as in our power analysis (below).

The \( r^2 \) test does not have a specific alternative hypothesis, meaning that a significant result suggests a power-law distribution is unlikely but does not indicate what other distribution might be correct. This is a characteristic of all widely used goodness-of-fit tests for statistical distributions [e.g., Filliben, 1975]. Consequently the statistical power (the probability of finding a significant \( r^2 \) test when the data are from a non-power-law distribution) must be evaluated for a variety of potential alternative distributions. We evaluated the power of the \( r^2 \) test against three alternate distributions with long upper tails (lognormal distribution, exponential distribution, chi square distribution). Alternative distributions were simulated 50,000 times for each sample size (10,000 for \( n = 1469 \)) and the power was determined as the number of \( r^2 \) values below the critical value divided by the number of simulated alternative distributions. The power of \( r^2 \) varies with sample size and alternative distribution (Table S1), but does not perform more poorly than tests commonly applied for other distributions [e.g., Razali and Wah, 2011].

**Table S1.** Critical values for significance and power analysis for the \( r^2 \) test.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( n = 19 )</th>
<th>( n = 29 )</th>
<th>( n = 114 )</th>
<th>( n = 210 )</th>
<th>( n = 1469 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical ( r^2 )</td>
<td>0.846</td>
<td>0.87</td>
<td>0.935</td>
<td>0.957</td>
<td>0.990</td>
</tr>
<tr>
<td>Power against lognormal (( \mu = 0, \sigma = 1 ))</td>
<td>0.415</td>
<td>0.574</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Power against exponential (( \lambda = 1 ))</td>
<td>0.871</td>
<td>0.974</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Power against chi square (df = 1)</td>
<td>0.944</td>
<td>0.994</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix B: Analysis of lake size data from Mars

We analyzed size data for open-basin lakes \( n = 210 \) on Mars [Fassett and Head, 2008]. The surface areas and elevations were determined by simulated flooding on a digital terrain model of the Martian surface [Fassett and Head, 2008]. The slope of the log-abundance log-size regression was \(-0.4\), which is not within the theoretical constraints of \( 0.5 \leq b \leq 1 \) (Fig. S1A). The \( r^2 \) value \( (r^2 = 0.93) \) was significantly lower than expected from data that conforms to a power-law distribution [Seekell and Pace, 2011]. The lake elevation distribution conformed to the normal distribution (Fig. S1B) and there was an order of magnitude more variability in lake elevation than in the Adirondack Mountains (mean elevation = 29m, standard deviation = 1385.3). The slope of the log-abundance log-size regression for Martian lakes near the mean elevation (mean = 29m, lakes selected from \( \pm 250 \) mean elevation, \( n = 29 \) lakes) was \(-0.539\), within the theoretical bounds (Fig. S1C). The \( r^2 \) value \( (r^2 = 0.945) \) was consistent with data that conform to a power-law distribution [Seekell and Pace, 2011]. The \( r^2 \) values for Figure S1A and S1C are similar but have imposing results because the critical value for significance changes with sample size [see Seekell and Pace, 2011 for details]. We could not use only lakes at the mean elevation as in the Adirondack Mountains because there was tremendous variability in the elevation distribution and multiple lakes did not fall exactly at the mean in this dataset. Selecting lakes near the mean value (mean elevation \( \pm 250 \)m) gave enough values for statistical analysis, but the range of elevations is very small relative to the total variability in the lake elevation distribution (Fig. S1B). The fractal dimension for the size-distribution of Mars lakes near the mean elevation is \( D = 1.08 \). Independently measured Hurst coefficients \( (H) \) for the Martian land surface vary considerably, but a common value for much of the surface is \( H = 0.72 \) [Orosei et al., 2003]. The relationship between the landscape Hurst coefficient and the expected fractal
Dimension of the lake size-distribution is $D = 2 - H$ [Goodchild, 1988]. Hence an expectation for Martian lakes near the mean elevation is $D = 1.28$. While there is a greater difference between these two fractal estimates for Mars than for the Adirondack Mountain data, our value derived from the lake size-distribution is reasonable because 1) the value is theoretically plausible, 2) there is inherent error in measuring lake sizes and elevations and uncertainty in the statistical estimation techniques, and 3) there is a wide range of Hurst coefficients reported for the Martian surface and the fractal dimension we report falls within the expected values based on this range of Hurst coefficients [Orosei et al., 2003]. For instance, another common Hurst coefficient estimate is $H = 0.84$, which is equivalent to an expectation of $D = 1.16$, similar to the value recovered from lakes near the mean elevation [Orosei et al., 2003].

We compared the fit of size distribution equations 1 and 2 for Mars away from the mean elevation by fitting the equation to lakes between 1750 and 2250 m. The sample size within this range of elevations was small ($n = 17$), but size-distribution equation 2 appears to have improved linear prediction ($r^2 = 0.98$) relative to size-distribution equation 1 ($r^2 = 0.908$) (Fig. S2). Further, the points appear to be more evenly distributed around the regression line for size-distribution equation 2 than for size-distribution equation 1. By the dual evaluation criteria described in the main text [see also Quandt, 1964], size-distribution equation 2 demonstrates improved fit relative to size-distribution equation 1.
Figure S1. A: Log-abundance log-size plot for ancient lake basins on Mars. The slope of the regression line is $-0.49$. The distribution of lake sizes departs significantly from the power-law distribution (the $r^2$ value, 0.93, is lower than what is expected from data from a power-law distribution, see Seekell and Pace [2011]). B: Histogram of elevations for Martian lakes. The mean elevation is 29m (standard deviation = 1385.3, range 10,440). The elevations are measured relative to a standardized reference surface [Smith et al., 2000]. The distribution of lake elevations conforms to a normal distribution. C: Log-abundance log-size plot for lakes near ($\pm$ 250m) the mean elevation. The slope of the regression line is $-0.539$. The $r^2$ value is 0.945. The high $r^2$ value suggests that the distribution of lake sizes near the mean elevation does not depart significantly from the power-law distribution.
Fig. S1

A. Abundance vs. Lake area (km²)
   - Slope = -0.40
   - $r^2 = 0.93$

B. Frequency vs. Elevation (m)
   - Mean = 29
   - Std. Dev. = 1385.3

C. Abundance vs. Lake area (km²)
   - Slope = -0.539
   - $r^2 = 0.945$
Figure S2. Observed versus fitted values for logarithm of abundance of Mars lakes ($n = 17$) away from the mean elevation based on log-abundance log size-regressions. Size-distribution equation 1 (Panel A) does not fit as well (i.e. the fit is less linear) as size-distribution equation 2 (Panel B).
References for Supplemental Material


