Limnol. Oceanogr., 56(1), 2011, 350–356

doi:10.4319/lo.2011.56.1.0350

Does the Pareto distribution adequately describe the size-distribution of lakes?

David A. Seekell* and Michael L. Pace

Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia

Abstract

When it comes to evaluating lakes at regional and global scales, a key need is accurate estimates of the abundance and size-distribution of lakes, which are usually described with the Pareto distribution. We demonstrate the considerable uncertainty that truncation in the lower tail of the Pareto distribution introduces into lake abundance estimates and the selection of the lake size-distribution. Truncation in the lower tail eliminates lakes below a certain size and is generally performed because small lakes are not accurately represented on maps. When simulated data are truncated to mimic available lake size data, non-Pareto distributions are visually and statistically indistinguishable from the Pareto distribution. The Pareto distribution may be one of many possible forms that mimic the global lake size-distribution in the upper tail, but the fit of the Pareto to the lower tail is uncertain, largely because the abundance of small lakes is uncertain. Some other potential size-distributions, such as the lognormal distribution, predict abundances of small lakes to be orders of magnitude lower than do the Pareto distribution predictions. Highly resolved regional lake size data for the Adirondack Mountains of New York and the Northern Highland Lake District of Wisconsin do not conform to the Pareto distribution. Lake sizes on Mars also do not conform to the Pareto. Uncertainty in the lake size-distribution seriously limits understanding of the significance of lakes as repositories of organic carbon as well as the calculation of global greenhouse gas emissions from these systems.

Recently, limnologists have become interested in developing more accurate estimates of the global distribution of inland aquatic systems (Lehner and Döll 2004; Downing et al. 2006). This interest is driven partly by the need to assess the importance of inland waters in processes such as the global carbon cycle (Cole et al. 2007; Battin et al. 2009; Tranvik et al. 2009) and partly by the need to more fully assess the numbers and sizes of aquatic systems, given global human pressures on inland waters (Wetzel 2001). This work has led to new findings, including that inland lakes and reservoirs cover a much greater portion of the earth's land surface ($\sim$ 3%) and that inland waters process substantial amounts of organic carbon, relative to previous estimates (Downing et al. 2006; Cole et al. 2007; Tranvik et al. 2009). At the global scale, Tranvik et al. (2009) estimated that land exports of carbon to inland waters are twice as high as land exports of carbon to the ocean. Most of this carbon is either subsequently exported to oceans (0.9 Pg yr^{-1}), is buried (0.6 Pg yr^{-1}), or is oxidized and evades into the atmosphere (at least 1.4 Pg yr^{-1}) (Tranvik et al. 2009). Lake sediments may contain as much as 820 Pg C (Cole et al. 2007). Globally, lakes are important methane sources, with greater emissions to the atmosphere than are provided by the world's oceans (Bastviken et al. 2004).

As with other global limnological analyses, these estimates of carbon processing and methane emission require an accurate estimate of the abundance and size-distribution of lakes (Tranvik et al. 2009). These estimates are particularly critical for assessing the abundance of small lakes, which may both contain and process large amounts of carbon per unit area (Kortelainen et al. 2004; Hanson et al. 2007; Telmer and Costa 2007). For instance, the smallest third of lakes in Finland, studied by Kortelainen et al. (2004), contained two thirds of the carbon stored by lakes in the region. Telmer and Costa (2007) found that including lakes measuring ≤ 0.1 km^2 increased the amount of carbon accounted for by lakes by as much as 30% on a per–unit area of landscape basis. Hanson et al. (2007) found that omitting small lakes from a regional study caused large biases when estimating mean dissolved inorganic carbon and dissolved organic carbon concentrations, even though total surface area in this study was dominated by large lakes.

The abundance of lakes is difficult to quantify because maps generally omit lakes below a certain size (Hamilton et al. 1992; Lehner and Döll 2004; Downing et al. 2006). A number of studies have used the Pareto distribution to estimate the global or regional abundances and surface areas of lakes (Lazzarino et al. 2009; Marotta et al. 2009; Tranvik et al. 2009). This approach fits a log-abundance log-size regression based on the largest lakes. Abundance is the number of lakes greater than or equal to a given size (Lehner and Döll 2004). For example, the abundance of lakes greater than or equal to 10 km^2 in the global lake size data of Lehner and Döll (2004) is 17,357. Parameters from the regression are then used to estimate the number of small, unobserved lakes using the Pareto distribution probability density function (Downing et al. 2006). For example, Fig. 1 displays the world's 17,357 lakes measuring $>$ 10 km^2 used by Downing et al. (2006) to estimate the global abundance of lakes. These data are well described by the Pareto distribution because of the excellent visual fit and high r^2 value. However, many distributions appear linear on a log–log plot when only the largest values are considered (Perline 2005; Gan et al. 2006). This apparently strong fit of the Pareto distribution to the data is a concern when estimating lake abundance because most lake size

* Corresponding author: das9xx@virginia.edu

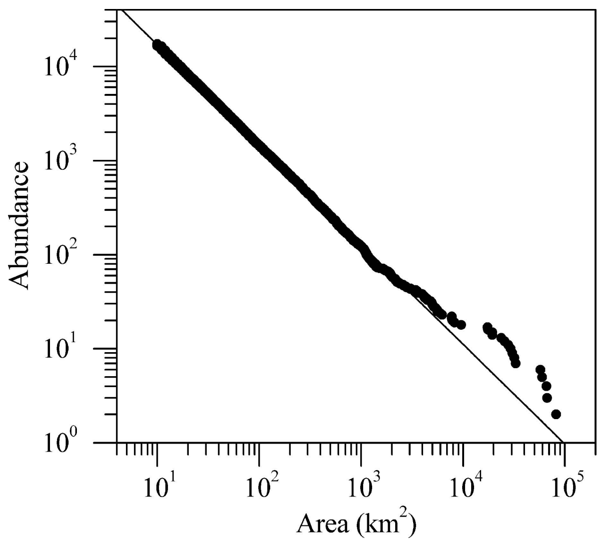

Fig. 1. Abundance-size data for the world's 17,357 lakes measuring greater than 10 km^2, from Lehner and Döll (2004), used by Downing et al. (2006) to estimate the global abundance of lakes. The ordinary least-squares regression slope is −1.06079 and the r^2 is 0.998. Based on the excellent linear fit and high r^2 value, it is reasonable to conclude that these data are well described by the Pareto distribution.

data are retrieved from maps that only include the largest lakes (Downing et al. 2006). For example, the data in Fig. 1 are truncated because lakes < 10 km^2 are excluded. Here we use both simulated and actual lake abundance-size data to evaluate the uncertainty in estimates of small lake abundance. When simulated size data are truncated to mimic available lake size data, non-Pareto distributions are visually and statistically indistinguishable from the Pareto distribution. Consequently, the lake size-distribution may be one of many possible forms that mimic the Pareto distribution in the upper tail. Some of these potential size-distributions, such as the lognormal distribution, predict small lake abundances that are orders of magnitude lower than the Pareto distribution prediction.

Methods

To test the potential goodness-of-fit to alternative distributions of small lakes, we simulated abundance-size data from the lognormal distribution (mean = 0, standard deviation = 1; when the data are natural log transformed) with no truncation and with high levels of truncation. We used truncations of 90%, 99%, and 99.9% to eliminate lakes from the lower tail of the distribution. The 90%, 99%, and 99.9% truncated data were created by generating 10,000; 100,000; and 1,000,000 random values, respectively, and then discarding all but the 1000 largest values for each sample (Perline 2005). We evaluated the data visually for linear fit on a log-log plot. To statistically evaluate the simulated truncated lognormal data relative to the Pareto distribution we employed the Kolmogorov–Smirnov test to determine if the truncated lognormal data were distinct from the Pareto fit.

We also evaluated the r^2 values from the log-abundance log-size regression for fit to the Pareto distribution. Data from the Pareto distribution are expected to fall along a straight line on a log–log plot. Thus, a low r^2 value represents a departure from linearity and is an indicator of departure from the Pareto distribution. Distributions other than the Pareto may have high r^2 values when regressed on a log–log plot (Gan et al. 2006). To discriminate against non-Pareto distributions that produce high r^2 values, we tested the departure from linearity statistically (Gaudoin et al. 2003). To evaluate the statistical significance of low r^2 values we calculated smoothed empirical percentage points using Monte Carlo simulation. We generated 10,000 random samples of $n = 1000$ from a Pareto distribution with shape and scale parameters equal to 1 (*see* Vidondo et al. [1997] or Downing et al. [2006] for details on the Pareto distribution). Gaudoin et al. (2003) has shown that r^2 is independent of the shape and scale parameter, and its distribution only depends on the sample size. For each sample we regressed the logarithm of abundance by the logarithm of size and calculated r^2. We then found the fifth quantile of the 10,000 simulated r^2 values. If the r^2 value is below the fifth quantile, then the data deviate from linearity enough that they are unlikely to be from a Pareto distribution (i.e., the probability of getting the r^2 value is less than 0.05, given that the data are actually from a Pareto distribution). We repeated the simulation five times and averaged the percentage points in order to have smoothed empirical percentage points because the distribution of r^2 is heavy-tailed. We provide smoothed empirical percentage points for other sample sizes in Table 1. Table 1 is used in the same way that a table for the *t*-distribution, *F*-distribution, or chi-square distribution (found, for example, at the end of an introductory statistics text) would be used.

To evaluate the uncertainty in lake abundance estimates associated with assuming different size-distributions, we generated 100,000 random values from a lognormal distribution (mean = 0, standard deviation = 2, when the data are natural log transformed; minimum value = 0.001). We fit a log-size log-abundance regression for the 1000 largest values (1% of the total). We also estimated the abundance of lakes based on the 1000 largest values using the Pareto distribution, according to Downing et al. (2006).

We created abundance-size plots for data from previously published high-resolution lake surveys in the Adirondack Mountains (New York) and the Northern Highland Lake District (Wisconsin) to evaluate departure from the Pareto distribution. The Adirondack Mountain lakes ($n = 1469$), with a minimum size of 0.001 km^2, were selected using a stratified random sampling scheme (Kretser et al. unpubl.). The Northern Highland Lake District data ($n = 7064$) have a minimum lake size of 0.0001 km^2 and represent a census of the region's lakes (Hanson et al. 2007). Both regions were subject to the geomorphic effects of the last glaciations. Most of the world's lakes are thought to occur on such glaciated terrains (Meybeck 1995), and, hence, these districts are representative of an important category of lakes. We also plotted data from Fassett and Head (2008) for open basin

Table 1. Smoothed empirical percentage points for r^2. The distribution of r^2 is dependent only on sample size (Gaudoin et al. 2003). We generated 100,000 random samples of size n from the Pareto distribution (location parameter = 1, scale parameter = 1) and calculated r^2 for the log-abundance log-size regression. We then tabulated the percentage points. Because the distribution is skewed, we repeated the simulation five times and averaged the percentage points to smooth them. If an r^2 of 0.87 was calculated for a sample size of 30, the Pareto distribution would be rejected, because the test statistic is below the critical value (0.871589) at a 0.05 level of significance. If an r^2 of 0.92 was calculated for a sample size of 30, the Pareto distribution would fail to be rejected, because the test statistic is greater than the critical value (0.871589) at the 0.05 level of significance.

	Level of significance										
n	0 (minimum)	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99	1 (maximum)
5	0.402663	0.654835	0.763113	0.80742	0.865335	0.922823	0.959632	0.979485	0.9869	0.995423	0.999694
10	0.543706	0.727838	0.806664	0.846633	0.898231	0.937344	0.962553	0.976918	0.98289	0.99038	0.99762
15	0.568744	0.74351	0.833144	0.871031	0.91692	0.948733	0.968129	0.979416	0.984239	0.990461	0.996461
20	0.589786	0.764707	0.850835	0.885803	0.927497	0.955596	0.972263	0.981921	0.98519	0.991292	0.996839
25	0.562291	0.771803	0.860483	0.896727	0.927852	0.960928	0.975511	0.983755	0.987203	0.991789	0.997043
30	0.586494	0.783155	0.871589	0.905365	0.941743	0.96476	0.977857	0.985119	0.988368	0.992507	0.996621
35	0.609485	0.796077	0.880617	0.912718	0.946875	0.967775	0.979614	0.986262	0.989115	0.992858	0.996924
40	0.580156	0.805851	0.887819	0.917887	0.950474	0.970261	0.981265	0.98739	0.990007	0.993431	0.997045
45	0.621022	0.812274	0.892736	0.922259	0.953538	0.972196	0.982533	0.988173	0.990646	0.993742	0.997073
50	0.594266	0.824738	0.898744	0.926139	0.956629	0.974034	0.983608	0.98883	0.991037	0.994034	0.99734
55	0.617768	0.830996	0.902898	0.930011	0.952815	0.975434	0.984529	0.988483	0.991613	0.994372	0.997553
60	0.627647	0.836168	0.908478	0.934044	0.961204	0.976789	0.985347	0.990064	0.992021	0.994662	0.997726
65	0.631666	0.841546	0.911335	0.936502	0.96272	0.977815	0.98611	0.990517	0.99238	0.994793	0.997679
70	0.657159	0.849528	0.914256	0.939011	0.964378	0.978834	0.986688	0.990922	0.992676	0.995071	0.997903
75	0.630534	0.857203	0.919211	0.941589	0.965837	0.979795	0.9873	0.991336	0.993009	0.995234	0.9978
80	0.642113	0.857607	0.920375	0.943647	0.967481	0.980846	0.987847	0.991632	0.99329	0.995432	0.997801
85	0.670023	0.862848	0.922433	0.945346	0.968129	0.981285	0.988281	0.992031	0.993548	0.995527	0.997926
90	0.673854	0.869481	0.925879	0.947076	0.969264	0.981896	0.988602	0.992197	0.993754	0.995767	0.998176
95	0.695583	0.873076	0.929116	0.949585	0.970609	0.982702	0.989101	0.992537	0.993964	0.99588	0.998409
100	0.699987	0.874597	0.929859	0.950341	0.971665	0.983314	0.989459	0.992777	0.994129	0.995977	0.998156
150	0.770416	0.898668	0.94491	0.961297	0.978079	0.987229	0.991929	0.994496	0.995524	0.996921	0.99852
200	0.763586	0.917998	0.954909	0.968289	0.981892	0.989426	0.993327	0.995373	0.996246	0.997386	0.998697
250	0.809559	0.928741	0.961026	0.972723	0.984426	0.990982	0.994315	0.996074	0.996802	0.997738	0.998934
300	0.824582	0.937231	0.966366	0.975983	0.986311	0.99202	0.99495	0.996522	0.997173	0.998032	0.99906
350	0.835515	0.944096	0.969667	0.978634	0.987701	0.992802	0.995474	0.996853	0.997428	0.998207	0.999058
400	0.841813	0.948963	0.97248	0.980474	0.988906	0.993505	0.995904	0.997162	0.997684	0.998362	0.999183
450	0.860647	0.954018	0.975058	0.982433	0.989833	0.994033	0.996209	0.997367	0.997854	0.99849	0.999244
500	0.821245	0.956484	0.976575	0.983341	0.990525	0.994447	0.99649	0.99734	0.997998	0.998585	0.999328
600	0.893725	0.96331	0.979877	0.985672	0.991719	0.995134	0.996925	0.997861	0.998249	0.998766	0.999385
700	0.900846	0.967127	0.982134	0.987361	0.992635	0.995672	0.997265	0.998094	0.998436	0.99889	0.999476
800	0.911311	0.970688	0.983935	0.988466	0.993317	0.996077	0.997507	0.998255	0.998574	0.998992	0.999493
900	0.915842	0.97353	0.985368	0.989479	0.993874	0.996402	0.9977	0.998398	0.998679	0.999074	0.999497
1000	0.924719	0.975308	0.986419	0.99024	0.994319	0.996673	0.99787	0.998512	0.998778	0.999135	0.999541
2000	0.950361	0.986129	0.992267	0.994419	0.996716	0.998048	0.998745	0.999118	0.999274	0.99948	0.99974
3000	0.967163	0.990286	0.994608	0.996111	0.997674	0.998587	0.999086	0.999353	0.999468	0.999621	0.9998
4000	0.973758	0.992511	0.995817	0.996929	0.998145	0.99887	0.999267	0.999483	0.999572	0.999696	0.999836
5000	0.9804	0.993959	0.996516	0.99747	0.998454	0.99906	0.999391	0.999568	0.999641	0.999742	0.999869

lakes on Mars (n = 210). Mars lake basins were identified over all longitudes and between 90°S and 60°N using digital terrain models, digital imagery, and the expert judgment of geomorphologists (Fassett and Head 2008). Lake surface area was estimated for individual Martian lakes by simulated flooding in the basin and by observing the size when water spilled into the lake outlet (Fassett and Head 2008).

Results

The simulated lognormal data appear Paretian in the upper tail and deviate considerably in the lower tail (Fig. 2). Despite being from a lognormal distribution, the three truncated data sets demonstrate excellent linearity and are visually indistinguishable from Paretian data (e.g., Fig. 1). Figure 2 demonstrates how the apparent visual fit (linearity) of the lognormal distribution increases as the degree of truncation increases. The simulated data shift to the right with increasing truncation because the values are drawn from a larger sample size. Rare, extreme values are more likely to appear in the larger sample sizes. Increasing levels of truncation on a single sample will also create a more linear appearance (Fig. 3). Even lognormal data with only the largest 10% of values appear linear (Paretian) on the log–log plot. For comparison, the global lake data used by Downing et al. (2006) contain only 17,357 of the world's estimated 304 million lakes.

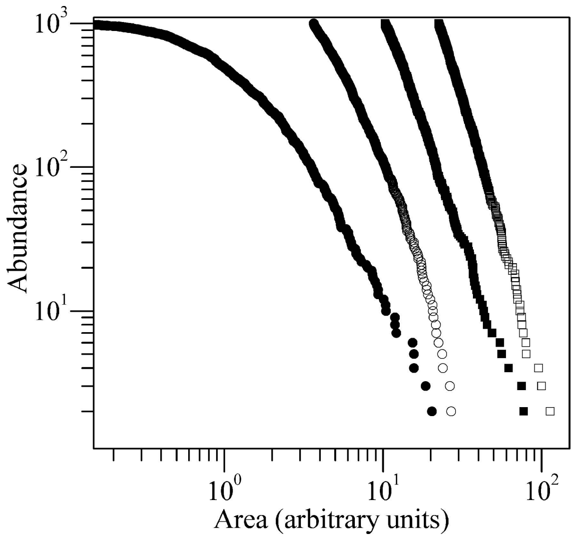

Fig. 2. Simulated data from a lognormal distribution (location parameter = 0, scale parameter = 1). The solid circles are 1000 points of random values. The open circles are the 1000 largest values from a sample of 10,000 (90% truncation). The solid squares are the 1000 largest values from a sample of 100,000 (99% truncation). The open squares are the 1000 largest values from a sample of 1,000,000 (99.9% truncation).

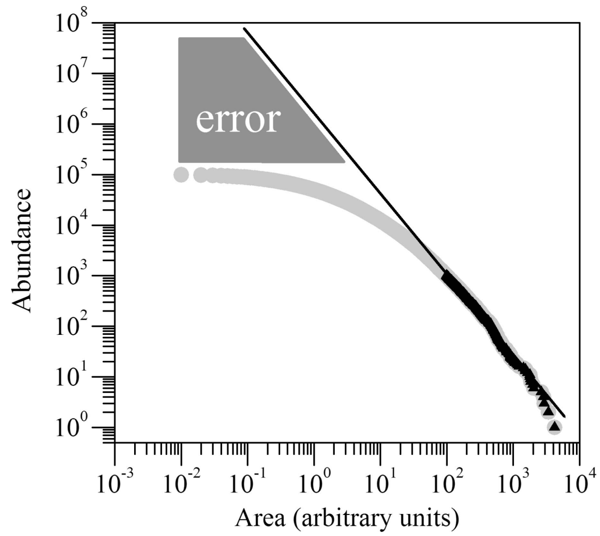

Fig. 3. Contrast between an extrapolated regression based on a Pareto distribution and the actual pattern based on a lognormal distribution for a hypothetical lake size-abundance relationship. The light gray circles are 100,000 simulated values from a lognormal distribution (minimum value = 0.001, location parameter = 0, scale parameter = 2). The superimposed black triangles denote the 1000 largest values. The black regression line is the log-abundance log-size fit for the 1000 largest values. The slope of the black regression line is -1.048 and the r^2 value is 0.997. For comparison, this slope and r^2 value are similar to the values identified by Downing et al. (2006) for the world's 17,357 largest lakes (slope = -1.0607, r^2 = 0.998). The 1000 largest values are indistinguishable from the Pareto distribution when analyzed separately from the rest of the lognormal data. Visual inspection of the regression line and data reveals that the Pareto distribution will grossly overestimate the number of small lakes, even though the large lakes appear to have an excellent fit to the Pareto. The portion of the figure representing the difference in the distribution is labeled "error" and is shaded dark gray.

The truncated lognormal data sets were not significantly different from the Pareto distribution (Table 2). Consequently, the truncated lognormal data are statistically indistinguishable from the Pareto distribution. While it is reasonable to assume that the data in Fig. 1 are well described by the Pareto distribution as a result of a high r^2, it cannot be excluded that small lakes not represented in lake size data follow another form. The lakes could easily follow a lognormal distribution, which has a similar upper tail to the Pareto distribution and would result in similar r^2 values. However, the lognormal distribution would predict considerably fewer small lakes and, thus, would change perspective on estimates of, for example, lake contributions to the global carbon cycle. To illustrate the magnitude of the possible error associated with using the wrong distribution, we plotted 100,000 randomly generated values from a lognormal distribution (in light gray in Fig. 3). The 1000 largest values (1% of the total) are denoted with superimposed black triangles (Fig. 3). The black regression line is the log-abundance log-size fit of the 1000 largest values. The slope of the black regression line is -1.048 and the r^2 value is 0.997, which is similar to the global lake-size relationship of Downing et al. (2006). In this case lake abundance estimates based on the Pareto distribution fit only to the largest values grossly overestimate the total abundance of lakes. With the Pareto distribution the 1000 largest lakes simulated are expected to be 5.88×10^{-7}% of the total number of lakes (recall the true value is 1%). This leads to an overestimate of small lake abundance by several orders of magnitude. Thus, there is the potential for considerable error if the true lake size-distribution is not Pareto because the lower tail is not represented correctly.

Lake abundance-size data from the Adirondack Mountains and the Northern Highland Lake District deviate considerably from the Pareto distribution (Fig. 4A,B). For

Table 2. Kolmogorov–Smirnov and r^2 test statistics for the Pareto distribution on data from the lognormal distribution with varying amounts of truncation (Fig. 2). An asterisk (*) indicates statistically significant ($p < 0.05$) deviation from the Pareto distribution. As the level of truncation increases, the two statistical tests are unable to reject lognormal data, meaning that these distributions are indistinguishable for the Pareto distribution.

Truncation	r^2	Conclusion	Kolmogorov–Smirnov	Conclusion
No truncation	0.833*	Reject Pareto	0.3599*	Reject Pareto
90% Truncation	0.992	Fail to reject Pareto	0.05509*	Reject Pareto
99% Truncation	0.995	Fail to reject Pareto	0.02944	Fail to reject Pareto
99.9% Truncation	0.989	Fail to reject Pareto	0.04043	Fail to reject Pareto

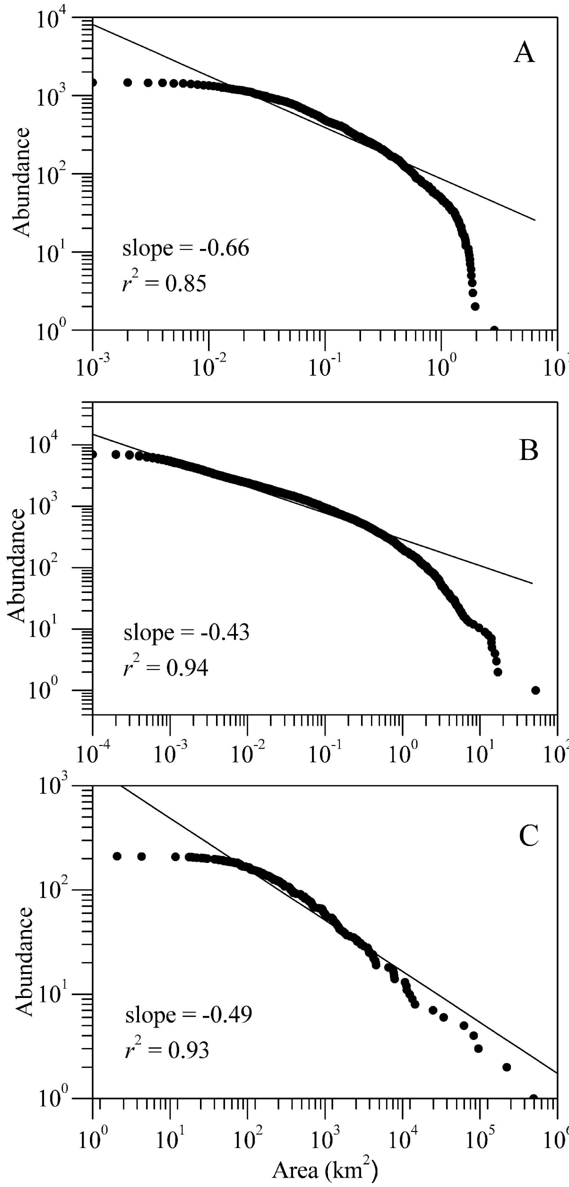

Fig. 4. Size-distribution of lakes in the (A) Adirondack Mountains, sampled by Kretser et al. (unpubl.) using a stratified random sampling scheme; (B) the Northern Highland Lake District lakes, from the lake census used in Hanson et al. (2007); and (C) open basin lakes identified by Fassett and Head (2008) on Mars. The slopes of the regressions and the r^2 values are shown on the figure. Note the difference in scales.

both regions, the abundance of small lakes is overestimated by the Pareto distribution and the abundance of medium-sized lakes is underestimated. The open basin lakes on Mars catalogued by Fassett and Head (2008) follow a similar pattern. Small Martian lakes are overestimated and medium lakes are underestimated using the Pareto distribution (Fig. 4C).

Discussion

A priority for regional and global limnological analyses should be to identify the true underlying size-distribution to ensure that the abundance of small lakes is not overestimated. Our recommendation contrasts with those of earlier studies (Hamilton et al. 1992; Lehner and Döll 2004; Downing et al. 2006) that expressed concern that the number of small lakes was underestimated. A major need of large-scale limnological studies is to ensure that samples are representative of the true population. In this respect we agree with the authors of prior studies, but the perspective we offer here is that recent studies may have overestimated the numbers of small lakes. Lakes of < 1 km^2 should receive increased attention in limnological studies because these systems (1) are likely the most abundant, even if distributions are more like a lognormal distribution than a Pareto distribution; (2) are currently thought to account for more surface area than large lakes (Downing et al. 2006); and (3) in some cases have poorly known or different processes relative to larger lakes (Fee and Hecky 1992; Pace et al. 2007).

The possibility that lake size-distributions more closely fit a lognormal distribution, or some distribution other than the Pareto distribution, is supported by the regional data sets for the Adirondack Mountains, the Northern Highland Lake District, and Mars. The first two studies involve relatively small lakes (*see* Methods). While these studies may underestimate the abundance of small lakes, the distributions are very flat in the lower tail. For example, in the case of the Adirondack Mountains, there is not a substantial increase in lake abundance for lakes below 0.02 km^2, which is well above the lower lake size cut-off for the study. Similarly, for the Northern Highland Lake District below about 0.001 km^2 lake abundance does not increase, as suggested by the Pareto. This threshold is an order of magnitude above the cut-off. Thus, the flattening is probably not the result of biased sampling of lakes that are as large as 20,000 m^2 and 1000 m^2 for the Adirondack Mountains and Northern Highland Lake District, respectively. In addition to the flat lower tails, the overall distributions for the Adirondack Mountains and Northern Highland Lake District are clearly curved on log–log plots, and this curvature is not simply the function of fewer small lakes. Downing et al. (2006) plotted several regional data sets in addition to the global size-distribution of large lakes. Some of these plots appear linear throughout a range of sizes. However, most of the highly resolved data sets plotted also appear to deviate from linearity for smaller lakes. Martian lakes appear more Paretian but also have a curved pattern on a log–log scale, with a flattening in the lower tail that is similar to that of lakes on Earth. In this case, the departure from the Pareto might be due to a bias in the sampling of small lakes.

A Pareto type II distribution, which is characterized by a flattened tail on a log-abundance log-size plot (Vidondo et al. 1997), is a possible alternative that might fit the data. However, the three lake districts we considered (Adirondacks, Northern Highlands, and Mars) were also significantly different from the Pareto type II distribution (results not shown). Furthermore, even if the Pareto II provided a good fit, the number of lakes in the lower tail (i.e., small lakes) would be considerably lower than a standard Pareto, as illustrated in Fig. 3.

The foregoing discussion about possible models to fit lake size distributions indicates that there remains no theoretical lake size-distribution. Wetzel (1990) originally postulated that the lake size-distribution might follow a Pareto distribution, but his argument was speculative and was not based on empirical or theoretical evidence (Downing 2009). Furthermore, there is no theoretical basis for a Pareto distribution of lake sizes in fractal geometry (Mandelbrot 1983). Derivation of a theoretical lake size-distribution as well as development of more accurate estimates of lake size and abundance (*see* below) should be a priority for global limnology.

Non-random sampling due to mapping processes that exclude or unreliably enumerate lakes below a size threshold almost certainly adversely affects estimates of lake abundance. New applications of remote sensing offer the means to solve this problem. Remote sensing has been applied to lakes in order to measure chlorophyll *a* (Tyler et al. 2006), water-column transparency (Chipman et al. 2004), and dissolved organic carbon (Winn et al. 2009). Some multispectral sensors, such as Ikonos, have both high spatial resolution (small pixel size) and high temporal resolution (the repeat time for taking images over a certain plot of land is short) and may have great utility in terms of observing small lakes (Folgo a Batista et al. 2003). Recent advances in object-oriented classification software allow for the identification of waterbodies based on both spectral signature and geometry without the need for time-consuming manual digitizing (Flanders et al. 2003). These advances should provide nearly complete enumeration of lakes in a region, as opposed to most current estimates, which are potentially inaccurate as a result of mapping omission and the types of statistical uncertainties we have presented here.

Advancing large-scale limnological analyses will also require adding additional variables to studies to provide more robust estimates of standing stocks and fluxes associated with lakes. For example, in the case of estimating CO_2 flux from lakes, additional metrics are needed to relate lake size to flux (Marotta et al. 2009). One error in flux estimates is introduced in the selection of the gas transfer velocity, in which it is often assumed that lakes of different sizes have the same average wind speed across their surface as the global average wind speed (Archer and Jacobson 2005; Marotta et al. 2009). With better information—for example, that provided by remote sensing—it should be feasible to scale gas transfer coefficients to attributes of a lake, such as size and position in the landscape.

Recent studies providing global estimates of lake abundance and carbon processing in inland waters have stimulated significant interest by evaluating and elevating the importance of inland waters in the global carbon cycle (Cole et al. 2007; Downing et al. 2008; Tranvik et al. 2009). These contributions, along with the analyses we present in this study, also point to the need for new approaches, methods, and analyses to generate estimates of inland water dynamics that go beyond relatively simple extrapolations.

Acknowledgments

We thank Jon Cole for discussions and advice, Steve Carpenter for providing data, and Kelly Hondula, William Lewis, and an anonymous reviewer for comments that greatly improved this article. This work was funded by the National Science Foundation (Division of Environmental Biology 0716869, Division of Environmental Biology 0917696); the University of Vermont Honors College; and the Department of Environmental Sciences, University of Virginia.

References

Archer, C. L., and M. Z. Jacobson. 2005. Evaluation of global wind power. J. Geophys. Res. **110:** D12110, doi:10.1029/2004JD005462

Battin, T. J., S. Luyssaert, L. A. Kaplan, A. K. Aufdenkampe, A. Richter, and L. J. Tranvik. 2009. The boundless carbon cycle. Nat. Geosci. **2:** 598–600, doi:10.1038/ngeo618

Chipman, J. W., T. M. Lillesand, J. E. Schmaltz, J. E. Leale, and M. J. Nordheim. 2004. Mapping lake water clarity with Landsat images in Wisconsin, U.S.A. Can. J. Remote Sens. **30:** 1–7.

Cole, J. J., and others. 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems **10:** 172–185, doi:10.1007/s10021-006-9013-8

Downing, J. A. 2009. Global limnology: Up-scaling aquatic services and processes to planet Earth. Verh. Int. Ver. Limnol. **30:** 1149–1166.

———, and others. 2006. The global abundance and size-distribution of lakes, ponds, and impoundments. Limnol. Oceanogr. **51:** 2388–2397, doi:10.4319/lo.2006.51.5.2388

———, and others. 2008. Sediment organic carbon burial in agriculturally eutrophic impoundments over the last century. Glob. Biogeochem. Cycles **22:** GB1018, doi:10.1029/2006GB002854

Fassett, C. I., and J. W. Head. 2008. Valley network-fed, open-basin lakes on Mars: Distribution and implications for Noachian surface and subsurface hydrology. Icarus **198:** 37–56, doi:10.1016/j.icarus.2008.06.016

Fee, E. J., and R. E. Hecky. 1992. Introduction to the Northwest Ontario Lake Size Series (NOLSS). Can. J. Fish. Aquat. Sci. **49:** 2434–2444, doi:10.1139/f92-269

Flanders, D., M. Hall-Beyer, and J. Pereverzoff. 2003. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. Can. J. Remote Sens. **29:** 441–452.

Folgo a Batista, M. T., A. C. Colaeo, S. G. Capelo, and J. M. Branco de Mascarenhas. 2003. Landsat/Ikonos applied to water quality monitoring in south of Portugal, p. 373–378. *In* M. Owe, G. D'Urso, and L. Toulios [eds.], Remote sensing for agriculture, ecosystems, and hydrology IV. International Society for Optical Engineering.

Gan, L., D. Li, and S. Song. 2006. Is the Zipf law spurious in explaining city-size distributions? Econ. Lett. **92:** 256–262, doi:10.1016/j.econlet.2006.03.004

Gaudoin, O., B. Yang, and M. Xie. 2003. A simple goodness-of-fit test for the power-law process, based on the Duane plot. IEEE T. Reliab. **52:** 69–74, doi:10.1109/TR.2002.805784

Hamilton, S. K., J. M. Melack, M. F. Goodchild, and W. M. Lewis. 1992. Estimation of the fractal dimension of terrain from lake size distributions, p. 145–163. *In* P. A. Carling and G. E. Petts [eds.], Lowland floodplain rivers: Geomorphological perspectives. Wiley.

Hanson, P. C., S. R. Carpenter, J. A. Cardille, M. T. Coe, and L. A. Winslow. 2007. Small lakes dominate a random sample of regional lake characteristics. Freshw. Biol. **52:** 814–822, doi:10.1111/j.1365-2427.2007.01730.x

Kortelainen, P., H. Pajunen, M. Rantakari, and M. Saarnisto. 2004. A large carbon pool and small sink in boreal Holocene lake sediments. Glob. Change Biol. **10:** 1648–1653, doi: 10.1111/j.1365-2486.2004.00848.x

Lazzarino, J. K., R. W. Bachmann, M. V. Hoyer, and D. E. Canfield. 2009. Carbon dioxide supersaturation in Florida lakes. Hydrobiologia **627:** 169–180, doi:10.1007/s10750-009-9723-y

Lehner, B., and P. Döll. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. **296:** 1–22, doi:10.1016/j.jhydrol.2004.03.028

Mandelbrot, B. B. 1983. The fractal geometry of nature. Freeman.

Marotta, H., C. M. Duarte, S. Sobek, and A. Enrich-Prast. 2009. Large CO_2 disequilibria in tropical lakes. Glob. Biogeochem. Cycles **23:** GB4022, doi:10.1029/2008GB003434

Meybeck, M. 1995. Global distribution of lakes, p. 1–35. *In* A. Lerman, D. M. Imboden, and J. R. Gat [eds.], Physics and chemistry of lakes. Springer-Verlag.

Pace, M. L., and others. 2007. Does terrestrial organic carbon subsidize the planktonic food web in a clear-water lake? Limnol. Oceanogr. **52:** 2177–2189, doi:10.4319/lo.2007.52.5.2177

Perline, R. 2005. Strong, weak and false inverse power laws. Stat. Sci. **20:** 68–88, doi:10.1214/088342304000000215

Telmer, K. H., and M. P. F. Costa. 2007. SAR-based estimates of the size distribution of lakes in Brazil and Canada: A tool for investigating carbon in lakes. Aquat. Conserv. **17:** 289–304, doi:10.1002/aqc.854

Tranvik, L. J., and others. 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnol. Oceanogr. **54:** 2298–2314.

Tyler, A. N., E. Svab, T. Preston, M. Présing, and W. A. Kovács. 2006. Remote sensing of the water quality of shallow lakes: A mixture modeling approach to quantifying phytoplankton in water characterized by high-suspended sediment. Int. J. Remote Sens. **27:** 1521–1537, doi:10.1080/01431160500419311

Vidondo, B., Y. T. Prairie, J. M. Blanco, and C. M. Duarte. 1997. Some aspects of the analysis of size spectra in aquatic ecology. Limnol. Oceanogr. **42:** 184–192, doi:10.4319/lo.1997.42.1.0184

Wetzel, R. G. 1990. Land-water interfaces: Metabolic and limnological regulators. Verh. Int. Ver. Limnol. **24:** 6–24.

———. 2001. Limnology: Lake and river ecosystems, 3rd ed. Elsevier.

Winn, N., C. E. Williamson, R. Abbitt, K. Rose, W. Renwick, M. Henry, and J. Saro. 2009. Modeling dissolved organic carbon in subalpine and alpine lakes with GIS and remote sensing. Landsc. Ecol. **24:** 807–816, doi:10.1007/s10980-009-9359-3

Associate editor: George W. Kling

Received: 24 March 2010
Accepted: 10 November 2010
Amended: 10 November 2010