AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2016GL071378

Key Points:

- Lake volume scales with lake area according to statistical topographic theory within and across diverse regions
- Total lake volume is 199,000 (196–202,000 95% CI) km^3 and is partitioned across size classes and given confidence bounds
- Total lake mean depth is 41.8 (41.2–42.4 95% CI) m, much lower than previous estimates

Supporting Information:

- Supporting Information S1

Correspondence to:
B. B. Cael,
snail@mit.edu

Citation:
Cael, B. B., A. J. Heathcote, and D. A. Seekell (2017), The volume and mean depth of Earth's lakes, *Geophys. Res. Lett.*, *44*, 209–218, doi:10.1002/2016GL071378.

Received 28 SEP 2016
Accepted 29 NOV 2016
Accepted article online 5 DEC 2016
Published online 13 JAN 2017

The volume and mean depth of Earth's lakes

B. B. Cael[1,2], A. J. Heathcote[3], and D. A. Seekell[4,5,6]

[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, [2]Department of Physical Oceanography, Woods Hole Oceanographic Institute, Woods Hole, Massachusetts, USA, [3]St. Croix Watershed Research Station, Science Museum of Minnesota, Marine on St. Croix, Minnesota, USA, [4]Climate Impacts Research Centre, Umeå University, Abisko, Sweden, [5]Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden, [6]Arctic Research Centre, Umeå University, Umeå, Sweden

Abstract Global lake volume estimates are scarce, highly variable, and poorly documented. We developed a rigorous method for estimating global lake depth and volume based on the Hurst coefficient of Earth's surface, which provides a mechanistic connection between lake area and volume. Volume-area scaling based on the Hurst coefficient is accurate and consistent when applied to lake data sets spanning diverse regions. We applied these relationships to a global lake area census to estimate global lake volume and depth. The volume of Earth's lakes is 199,000 km^3 (95% confidence interval 196,000–202,000 km^3). This volume is in the range of historical estimates (166,000–280,000 km^3), but the overall mean depth of 41.8 m (95% CI 41.2–42.4 m) is significantly lower than previous estimates (62–151 m). These results highlight and constrain the relative scarcity of lake waters in the hydrosphere and have implications for the role of lakes in global biogeochemical cycles.

1. Introduction

Accurate estimates of the volume of lakes are of fundamental interest in limnology. Additionally, volume estimates provide for the calculation of mean depth (lake volume/lake area), a correlate of many key lake ecosystem functions [*Vollenweider*, 1975; *Canfield and Bachman*, 1981; *Bastviken et al.*, 2008]. Despite this, there are few estimates of the global lake volume; historic estimates are highly variable, typically poorly documented, often without any data, methods, or references to other sources (Text S1 and Tables S1 and S2 in the supporting information). Hence, there is need to develop robust and well-documented estimates of Earth's lake volume.

Lake volume measurements require detailed bathymetric surveys, which are expensive and time consuming to conduct [*Hollister and Milstead*, 2010]. A growing number of regional reports use lake area and lakeside topography (e.g., maximum slope in a 50 m wide buffer around the lake shoreline) to predict individual lake depth and volume [e.g., *Sobek et al.*, 2011; *Heathcote et al.*, 2015]. This information is readily derived from maps, and these approaches have the potential to greatly expand the number of lakes for which volume and lake depth is known. However, the prediction accuracy for these approaches is highly variable. Additionally, the topographic characteristics identified as the best correlates for depth and volume vary regionally. This suggests that these methods are region specific and not generalizable to the global scale [*Oliver et al.*, 2016]. Alternate approaches are therefore needed to make global volume estimates.

The accuracy of lake surface area estimates has improved substantially over the past 15 years, mostly due to improved accounting of small lakes that were traditionally omitted from maps and lake registers [*Downing et al.*, 2006; *Seekell and Pace*, 2011; *Verpoorter et al.*, 2014]. Ideally, improvements in lake area databases could be leveraged to improve lake volume estimates. Here we develop a mechanistic method for estimating the global lake volume and mean depth with confidence bounds based on lake area data and mathematical theories for the characteristics of self-affine surfaces. Our approach is general and represents an important step forward in connecting theory with data in evaluating global-scale lake characteristics.

2. Methods

2.1. Theory

Many geophysical problems have been addressed by characterizing Earth's topography as a self-affine surface [*Goodchild*, 1988; *Ouchi and Matsushita*, 1992; *Cox and Wang*, 1993; *Seekell et al.*, 2013]. Self-affine surfaces are null models; they represent an idealized topography without the influence of scale-dependent geomorphic processes [*Goodchild*, 1988]. On self-affine surfaces, the relationship between topographic variations and horizontal scale is characterized by the Hurst exponent H [*Dodds and Rothman*, 2000], such that for a surface $z(x, y)$ any rescaling by a coefficient b conforms, in a statistical sense, to

$$b^{-H}z(b\,x, b\,y) = z(x, y)$$

This indicates that for a 2-D shape defined by $\ell := \{x, y|x, y \in \ell\}$, rescaling ℓ by $b^{1/2}$ will rescale its area $a(\ell)$ by b. The mean elevation (or mean depth) $\bar{z}(\ell)$ will rescale by $b^{H/2}$, which is equivalent to $\bar{z}(\ell) \sim a^{H/2}$. Mean depth relates lake volume to lake area (i.e., $v = \bar{z}a$); and therefore, the volume v contained by $\bar{z}(\ell)$ scales as $v(\ell) = a(\ell)\bar{z} \sim a^{1+H/2}$ [*Carpenter*, 1983; *Wetzel and Likens*, 2000]. In essence, for self-affine surfaces, there is a statistical relationship between horizontal and vertical scales, characterized by H; this relationship yields a statistical volume-area relationship for lakes passively embedded in a self-affine surface (see Figure 1).

Previously reported measurements of Earth's surface topography indicate that, for scales relevant to lakes (i.e., above 10^2 m), self-affine surfaces adequately characterize the first-order characteristics of landscapes, and that the Hurst coefficient at these scales is around $H = 0.4 \pm {\sim}0.1$ [*Mark and Aronson*, 1984; *Dodds and Rothman*, 2000; *McClean and Evans*, 2000; *Renard et al.*, 2013]. This indicates that lake volume should scale with area as

$$v \sim a^{1.2\pm0.05}$$

This scaling is statistical rather than exact. In other words, it is suitable only for predicting the volume-area relationship for *collections* of lakes, not for predicting individual lake volumes from their areas. This is in contrast to data-driven approaches in which the lakeside topography is reflected in the bathymetry of individual lakes. The assumptions of this approach are (i) the surface of the Earth is self-affine at relevant scales for lakes, with Hurst exponent $H = 0.4 \pm {\sim}0.1$, and (ii) lakes are objects that fill in depressions of this surface. The scaling relationship here represents both a geometric explanation for variation in lake volumes and a testable hypothesis, which can be evaluated using empirical scaling relationships [cf. *Seekell et al.*, 2013].

2.2. Empirical Test of Scaling Relationships

We tested our volume-area scaling prediction based on a collection of lake data sets spanning diverse regions. Specifically, we used volume measurements calculated from bathymetric surveys from the Adirondack Mountains (New York, USA), Wisconsin (USA), Quebec (Canada), and Sweden. Collectively, these surveys cover 22 biophysical regions (Table S3). The survey data were developed using standard limnological methods (Table S3). We tested the theoretical scaling relationship outlined above by fitting log-volume log-area regressions using the formula $\log v = \zeta \log a + \log \kappa$ (where ζ is the volume-area scaling exponent and κ is a proportionality coefficient) to these databases and comparing the empirical values to the predicted value. We developed 95% confidence limits for these coefficients by bootstrapping. We then combined these data sets and evaluate the overall volume-area scaling exponent.

We also evaluated volume data from the continental USA based on the Environmental Protection Agency's (EPA) Eastern and Western Lakes Survey. These surveys have broad geographic coverage but used nonstandard methods (Table S3). Because of this, we are less confident in the relative quality of these measurements. Therefore, we evaluated the fit of these data to the empirical scaling coefficient calibrated based on the combined data from the four standard surveys described above. We treated data on Earth's largest lakes the same way. These large lakes have global coverage, but the reported volumes are sometimes variable and poorly documented. We treat the EPA and large lakes data differently because we have less confidence in the quality of their measurements but include them here to ensure that we evaluate the generality of scaling relationships across the full size spectrum of lake areas and across scales and regions (i.e., from a single geographic region to the continental scale). Collectively, these six data sets include lakes formed by glacial, tectonic, and other processes, located both north and south of the last glacial maximum (Table S3).

2.3. Development of Global Lake Volume Estimates and Confidence Intervals

We use the characteristics of the above volume-area scaling relationship from the combined regional data sets to estimate total lake volume based on surface areas from a global lake census. We also estimate

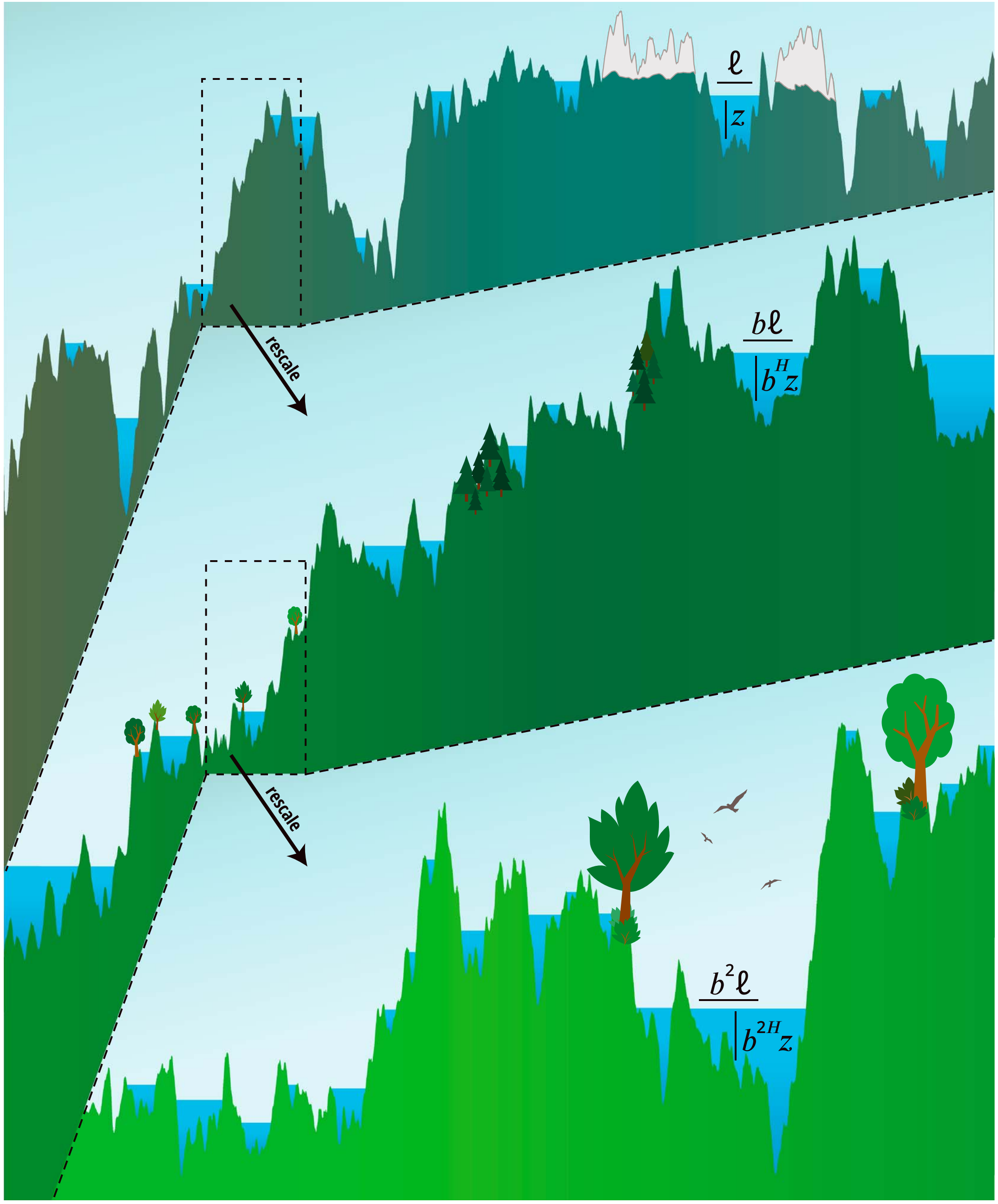

Figure 1. Schematic of statistical topographic scaling giving rise to lake volume-area scaling. Pictured is a section of a self-affine landscape with $H = 0.4$. As one rescales, or "zooms in" by a factor of b horizontally and b^H vertically, the landscape is statistically identical. If lakes are passively embedded on this landscape, then lake mean depth should scale with lake area according to $v \sim a^{1.2 \pm 0.05}$.

confidence intervals for the global volume estimate, incorporating the uncertainty in the pair (ζ, κ) via bootstrap resampling, and uncertainty due to the variability of lake volumes observed for each specific lake area via a Monte Carlo procedure. The Monte Carlo procedure directly estimates the influence of fluctuations around the scaling relationship, avoiding the need for a transformation bias correction [*Beauchamp and Olson*, 1973; *Smith*, 1993].

For our analysis we used the global lake census developed and described in detail in *Verpoorter et al*. [2014]. Briefly, these authors applied an automated lake extraction algorithm to cloudless, high-resolution satellite imagery taken around the year 2000. The data have previously been validated and were found to be highly accurate both in terms of numbers of lake recorded and their surface areas [*Verpoorter et al*., 2012]. The census includes artificial water bodies. For our scaling analysis, we excluded the 20 largest lakes because their volumes are well defined, to maximally constrain volume estimates. We applied the scaling relationships to the remaining lakes and then subsequently added the volumes for the largest lakes back to the total estimated from the scaling analysis [*Ryanzhin*, 2005]. We included lakes $\geq 0.01\ \mathrm{km}^2$ in our analysis, 28 million of which are recorded in *Verpoorter et al*. [2014]. Because the *Verpoorter et al*. [2014] lake area database is known to be highly accurate, as are volume estimates of the largest 20 lakes, we assume any measurement error in lake area has comparatively small influence on the total volume estimate and we do not include this in the development of the confidence intervals.

We performed 10,000 estimates of total lake volume. For each iteration, we estimate (ζ, κ) simultaneously via bootstrap resampling of the scaling relationship for the combined regional data sets (ζ and κ cannot be estimated independently because they are inversely correlated). We then apply this scaling relationship to each lake area measurement (m^2) from the global lake census, along with a multiplicative error term consistent with the distribution of residuals about the scaling relationship for the combined regional data sets. That is, fluctuations around the scaling relationship are directly incorporated by a multiplicative error term, which is a random variable normally distributed with mean zero in log space and with standard deviation ε given by the root-mean-square error (RMSE) of the residuals of the scaling relationship. We then sum over all lake volume estimates (including the 20 lakes excluded from the bootstrapping/Monte Carlo procedure) to obtain a global lake volume estimate. The total formula for this volume estimate is then given by

$$V := \sum v = \sum_a 10^{\kappa+\varepsilon} a^{\zeta}$$

where V is total lake volume and all other symbols are as above. We obtained an estimate and confidence intervals by calculating the corresponding quantiles from the distribution of the 10,000 total volume estimates. From these we then estimated mean depth (total lake volume divided by the total lake area) and average volume (total lake volume divided by total lake abundance).

To summarize these methods, we found that lake volume-area data vary in a specific way about a particular area-volume relationship. We characterized this relationship and variation, then randomly simulated lake volumes from lake area data such that they conformed to both. We then estimated the total volume by summing these simulated volumes and repeated the procedure many times to quantify the uncertainty produced by the variation about the relationship as well as the uncertainty in the scaling relationship itself.

The bootstrap procedure and the Monte Carlo procedure are applied together to account for different sources of uncertainty. The bootstrap procedure accounts for uncertainty in ζ and κ (analogous to not knowing whether a coin is fair or slightly biased). Even if the slope and intercept were known exactly, because individual lake volumes fluctuate around this scaling relationship, summing all lake volumes produces an additional uncertainty (analogous to counting the number of heads when a fair coin is flipped 100 times). The procedure used herein accounts for both types of uncertainty, by simulating individual lake volumes from individual lake areas using (i) a distribution of (ζ, κ) drawn from data and (ii) a random coefficient whose statistics are drawn from data. To test whether either source of uncertainty may be negligible, we repeat the estimation procedure above twice, accounting for only one source of uncertainty each time. To test only for slope-intercept uncertainty, we replace the random coefficient 10^{ε} with a transformation bias correction $\mathrm{TBC} = \mathrm{mean}(10^{\varepsilon}) = 1.254$ [*Beauchamp and Olson*, 1973; *Smith*, 1993]. To test only for uncertainty from summing random variables, we fix the slope-intercept pair as the median of the bootstrap distribution of slope-intercept pairs. We then compare the width of these confidence intervals to the width of the confidence interval resulting from the original estimation procedure.

3. Results

3.1. Scaling Relationships

The overall scaling relationship based on the lake bathymetry surveys was

$$\log_{10}(v) = 1.204\log_{10}(a) - 0.629 + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0,\ \sigma)$$

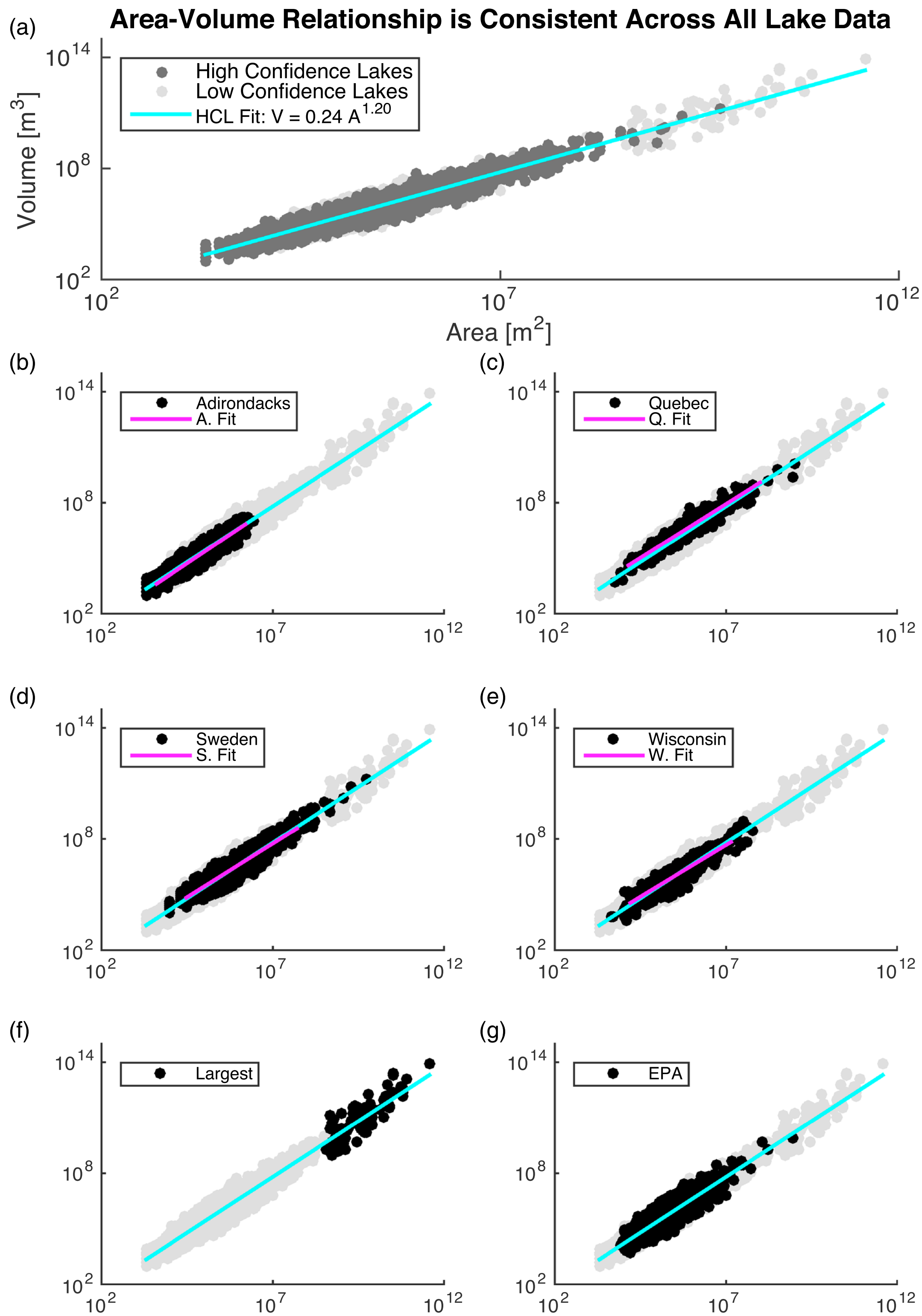

Figure 2. Volume-area scaling is consistent with Earth's Hurst exponent across all lake data and within high-confidence lake data sets. (a) Cyan line is the OLS linear regression model of $\log_{10}$(volume) on $\log_{10}$(area) for high-confidence lakes, with a scaling estimate of 1.204, corresponding to a Hurst exponent of $H = 0.41$. High-confidence lakes are plotted in dark grey, and low-confidence lakes are plotted in light grey. (b–e) Each subplot shows one high-confidence lake data set plotted in black and corresponding OLS linear regression model in magenta; compare each to cyan line. (f and g) Black points are low-confidence data sets. Residuals of high-confidence regression model prediction of low-confidence data sets are normally distributed with mean indistinguishable from zero. See supporting information. GIF of this figure is available at http://cael.space.

Table 1. Regional Scaling Exponents and Bootstrap 95% Confidence Limits, and Proportionality Coefficient[a]

Region	No. Lakes	Scaling Exponent ζ	95% CI	Coefficient κ	Mean Residual	RMSE
Combined	5516	1.204	(1.193, 1.215)	0.533	0	0.292
Adirondack	1469	1.239	(1.214, 1.263)	0.425	0.0579	0.296
Quebec	424	1.165	(1.124, 1.208)	0.758	−0.1784	0.285
Sweden	2269	1.16	(1.142, 1.177)	0.694	−0.0081	0.288
Wisconsin	1354	1.118	(1.091, 1.146)	0.855	0.0066	0.272
EPA	1860	---	---	---	−0.0059	0.395
Large lakes	77	---	---	---	−0.0041	0.603

[a]RMSE and mean residual are relative to the overall scaling relationship. The "Combined" data set is the aggregate of the higher confidence (Adirondack, Quebec, Sweden, and Wisconsin) data sets. Scaling estimates were not computed for lower confidence data sets (EPA and Large Lakes), but rather the scaling relationship from the Combined data set was used to compute the residuals; see Table S3 for descriptions of data sets.

where $\sigma = 0.292$ is the RMSE of the residuals. The bootstrapped 95% confidence interval for the scaling coefficient is $\zeta = 1.193$–1.215. Hence, the overall scaling relationship was consistent with values expected based on previously reported values for Earth's Hurst coefficient, and the 95% confidence limits are within plausible ranges based on variability in reports of Earth's Hurst coefficient.

Figure 2 displays the scaling relationships for the combined data (a) and the regional data sets (b–g). For the overall relationship, the calibration data are given as dark circles. The lower confidence lakes are drawn in grey. The regional fits were consistent with the overall fit, with scaling exponents within 0.1 of the overall scaling relationship (Table 1 and Figure 2). In each of the lower panels, the data from the individual regions are drawn black and the rest of the data are given in grey. Each region fits well within the overall data cloud. The RMSE and mean residual for each region relative to the overall scaling relationship is given in Table 1. The RMSE is consistent across the standard bathymetric surveys. Bootstrapped confidence intervals mostly overlapped with the expected range of coefficients (Table 1). These results emphasize the generality of the overall scaling relationship (Table 1). The two lower quality data sets also fit well relative to the overall scaling relationship, having mean residuals close to zero, albeit with somewhat larger RMSE values (Table 1). The larger residual variance is not surprising giving the sampling uncertainties in these data, but the fact that the overall scaling relationship still fits emphasizes the robustness and generality of the scaling result across the entire range of lake areas.

Lake area and residuals from the scaling relationship are uncorrelated ($p > 0.5$), and residuals from the scaling relationships are normally distributed (see Table S4). This indicates that fluctuations in the volume-area scaling relationship are multiplicative and uniform across all lake sizes. In other words, for a collection of lakes with the same surface area a there should be a lognormal distribution of volumes, and the variance of that distribution in log space is independent of a. The total volume estimate is the sum across these lognormal distributions. This also justifies the use of ordinary least squares as a slope estimator, as well as the approximation of residuals as normal in log space in volume estimation. These results are consistent across the regional and the combined data sets.

3.2. Global Estimates

Our estimate for the volume of Earth's lakes is 199,000 km^3 (95% confidence interval 196,000–202,000 km^3). This estimate is on the lower end of previous estimates, which average 210,000 km^3 (Table S2). This volume implies an overall mean depth (defined above as total volume divided by total area) of 41.8 m, which is substantially lower than previous reports (Table S2). This low mean depth is a consequence of the relatively lower volume estimate and the relatively higher surface area recorded in the lake census. This finding reflects the influence of small lakes, which were undercounted in earlier estimates [*Downing et al.*, 2006].

In addition to estimating total volume, we used the above procedure to estimate the volume contained in logarithmic area size classes. In Table 2, we provide abundance, total area, total volume (with confidence bounds), mean depth, and average volume of lakes in each area size class. Most of the volume of Earth's lakes is held in just a few large lakes, consistent with previous reports [e.g., *Herdendorf*, 1982; *Ryanzhin*, 2005]; the challenge of accurately estimating total volume resides in estimating the contribution of the millions of remaining lakes. The fraction of total volume declines with logarithmic area size class, though even small

Table 2. Abundance, Total Area, Total Volume, Mean Depth, and Average Volume of Lakes in Each Order of Magnitude Area Class[a]

Area Size Class	Abundance	Total Area (km^2)	Total Volume (km^3)	Mean Depth (m)	Average Volume (m^3)
10^4–10^5 m^2	23,725,071	683,000	1,780 (1730, 1840)	2.60 (2.53, 2.69)	750 (729, 776) $\times 10^2$
10^5–10^6 m^2	3,813,612	995,000	4,050 (3990, 4120)	4.07 (4.01, 4.14)	106 (105, 108) $\times 10^4$
10^6–10^7 m^2	331,452	793,000	5,080 (4940, 5240)	6.41 (6.23, 6.61)	153 (149, 158) $\times 10^5$
10^7–10^8 m^2	24,332	611,000	6,340 (6030, 6690)	10.4 (9.87, 10.9)	261 (248, 275) $\times 10^6$
10^8–10^9 m^2	1,948	489,000	8,110 (7520, 8760)	16.6 (15.4, 17.9)	416 (386, 450) $\times 10^7$
10^9–10^{10} m^2	211	537,000	14,200 (12900, 15700)	26.6 (24.0, 29.2)	673 (611, 744) $\times 10^8$
$> 10^{10}$ m^2	20	1,020,000	160,000	157	800×10^{10}
Total	27,896,646	5,130,000	$199{,}000 \pm 3{,}000$	41.8 (41.2, 42.4)	713 (703, 724) $\times 10^4$

[a]The 95% confidence intervals are reported in parentheses. Number and total area are computed directly from the lake area census used herein [*Verpoorter et al.*, 2014]; total volume and confidence bounds are estimated from quantiles of the bootstrapping and Monte Carlo procedure described in the text. Data for lakes $> 10^{10}$ m^2 are taken from *Ryanzhin* [2005].

lakes contain a nonnegligible portion of total volume because of the rapid increase of lake abundance with decreasing size; lakes with surface area 10^5–10^6 m^2 still contain 2% of total volume. This significant contribution by even small lakes is attributable to the improvement of detection of small lakes in the lake census used for this estimation as compared to previous global lake area censuses. The monotonic decay in volume contribution, along with the negligible contribution of lakes 10^4 – 10^5 m^2 to the total volume, indicates that lakes $< 10^4$ m^2 would not contribute significantly if we had included them in our analysis.

We also note that though the largest lakes dominate total volume, excluding fewer of the largest lakes from our scaling estimation procedure did not change our results. When the same procedure as above is repeated, but only excluding and subsequently adding volumes of the Caspian Sea, Lake Baikal, Lake Victoria, Lake Tanganyika, and the North American Great Lakes, our total volume estimate's 95% confidence interval is 193,000-203,000 km^3, which contains the confidence interval when the 20 largest lake volumes have been fixed. This demonstrates the estimate is not sensitive to the number of lake volumes that are fixed by direct measurement and emphasizes that the method can be applied to any combination of lakes with known and unknown volumes as long as their areas are known.

We found that 10^4 iterations of the estimation procedure are sufficient to ensure convergence of the estimate and confidence intervals, as quartiles from 2500 estimate subsets of the total distribution agree with the quartiles of the total distribution to three significant digits. Repeating the estimation with a transformation bias correction replacing the Monte Carlo procedure resulted in the same estimate with narrower confidence bounds (197,000–201,000 km^3); repeating the estimation with the Monte Carlo procedure and a fixed slope-intercept pair $(\zeta, \kappa) = (1.204, -0.629)$ also resulted in the same estimate with narrower confidence bounds (also 197,000–201,000 km^3). This suggests both sources of uncertainty are important in estimating confidence bounds, and that the total volume estimate of 199,000 km^3 is robust to these changes in the estimation procedure.

4. Discussion

How many lakes are there and how much water do they hold? This is one of the most fundamental questions in limnology. Estimates of the total lake surface area have increased over the past two decades, and recent studies have constrained this value (Figure S1 and Table S1). However, formal estimates of lake volume and depth have been elusive and there has been no change in these estimates, even as surface areas have been continually revised (Figure S1). Our estimate of volume of Earth's lakes is on the low end of the range of previous estimates, and the confidence intervals indicate that this result is well constrained. To our knowledge, this is the first analysis that has given a mechanistically based estimate of the global lake volume, and the first that has quantified uncertainty in this estimate in a statistically rigorous fashion. This estimate better constrains the influence of small lakes on total volume than past estimates, and we provide simultaneously abundance, total area, and total volume estimates for each area size class for lakes globally.

These improvements in the size partitioning, uncertainty quantification, and theoretical grounding of total lake volume alone represent a major step forward in understanding the distribution of lake water on the

Earth's surface; however, the substantially lower mean depth estimate of 41.8 m may also have significant implications for the role of lakes in global geochemical cycles. Lake depth has been used for decades as the best estimator of the sensitivity of lakes to nutrient loading [*Vollenweider*, 1975] and has similarly proven to be a reliable predictor for primary production and lake transparency [*Canfield and Bachman*, 1981]. The existence of more shallow lakes indicates a greater percentage of lakes could be susceptible to eutrophication via nutrient loading with increasing primary production and deceasing transparency. More recently, lake depth has been shown to be a key parameter for predicting lake CH_4 emissions due to its constraint on the area where ebullitive fluxes (the direct transport of methane from lake sediments to the atmosphere via bubbles) can occur [*Bastviken et al*., 2008]. Our result of a lower predicted global mean depth would indicate a greater relative contribution of ebullition to lacustrine methane fluxes emphasizing the importance of measuring this particular fraction in addition to diffusive flux. These examples illustrate both the importance of a well-constrained estimate of global mean depth and the potentially significant effect of lowering that estimate as lakes are increasingly seen as globally important processors of carbon while at the same time becoming increasingly sensitive to global trends in eutrophication.

Because our approach is theoretically driven, it is general in the sense that it is applicable to any self-affine surface. This generality is important for two reasons. First, data-driven approaches are often region specific and therefore not appropriate for making global-scale estimates [cf. *Oliver et al*., 2016]. Second, many bathymetric surveys are not based on representative samples, a common problem when evaluating lake characteristics [*Hanson et al*., 2007; *Wagner et al*., 2008]. Sampling bias for lake bathymetry can be related to geography, size, accessibility, or level of public interest. For example, volumes in Sweden are disproportionately known for southern versus northern lakes. This is because acidification strongly impacted lakes in southern Sweden, and volumes were measured to determine the appropriate amount of lime to add as a counter measure [*Håkanson and Karlsson*, 1984]. Lakes in northern Sweden were not strongly impacted by acid rain and hence less is known about lake volumes. For the EPA Eastern Lakes survey, biases exist because only lakes $> 0.04\ \mathrm{km}^2$ were sampled [*Landers et al*., 1988]. This excludes the majority of the United States' lakes [*Hanson et al*., 2007; *McDonald et al*., 2012]. The data this study based on are not immune to these potential biases, but the consistency between theoretical predictions and diverse data sets suggests that the potential impact of any bias is minimized in our analysis.

The generality of our approach can be further illustrated by examining reports from other self-affine surfaces [e.g., *Seekell et al*., 2013]. For example, the Hurst coefficient for Mars is $H = 0.7$ (standard deviation 0.19), which gives the expected scaling relationship $v \sim a^{1.35 \pm 0.095}$. *Fassett and Head* [2008] used topographic data and satellite imagery from Mars to estimate the volume of 210 Noachian-aged drainage lakes. These authors reported a volume-area scaling exponent of $\zeta = 1.31$ (RMSE = 0.35), which is consistent with the independently measured Hurst coefficients. These literature results emphasize the generality of using the Hurst coefficient as a mechanistic connection between lake area and volume is not restricted to the current conditions of Earth's surface. The generality of the method described herein means it can be straightforwardly used to estimate total volume for specific regions and using other global lake area censuses. Consistent with the general applicability discussed above, within the lakes used in our empirical analyses, glacial and nonglacial lakes scale similarly, in accordance with Earth's Hurst coefficient ($H = 0.4$).

The precision of our global lake volume estimate relies strongly on the quality of surface area measurements in a global lake census. The global lake census our volume estimate is based on is thought to be the best available, but we realize the area estimates from this data set fall at the upper end of the range of historic estimates (Table S1). Even if we assume the *Verpoorter et al*. [2014] lake census overestimated global lake area our volume estimates would not be strongly impacted because scaling coefficients indicate small lakes, where most uncertainty exists in lake abundance and area, contain only a small fraction of total lake volume. However, because these small lakes have a larger relative contribution to total lake surface area than total lake volume (Table 2), such an overestimation could impact the overall mean depth of lakes. To evaluate the robustness of our volume estimate relative to the underlying lake database, we applied our volume scaling relationship to the Global Lake and Wetlands database compiled by *Lehner and Döll* [2004]. This database contains records for $2.43 \times 10^6\ \mathrm{km}^2$ of lake surface area, which is on the low end of the range of contemporary estimates (Table S1). Specifically, the database is thought to underestimate the abundance of lakes $< 10\ \mathrm{km}^2$ and is known to significantly underestimate lakes $< 1\ \mathrm{km}^2$ [*Lehner and Döll*, 2004; *Downing et al*., 2006]. If we applyour scaling relationship to these data, we estimate a global lake volume of 184,000 km^3 (95% confidence interval 181,000–186,000 km^3) and an overall mean depth of 75.7 m. While the specific volume estimate between data sets is slightly lower, our overall conclusion of a lower lake volume and mean depth than in historic estimates is robust to the lake census underlying the estimate and to the underestimation of small lakes.

Because many key ecosystem characteristics scale with either lake area or volume, the asymmetry between size and abundance is a key constraint on global patterns in lake ecology [*Cael and Seekell*, 2016]. Large lakes contribute most to the lake volume, and this contrasts patterns for lake abundance, for which small lakes contribute most, and patterns for surface area, for which medium-sized lakes contribute most [*Hanson et al.*, 2007; *Verpoorter et al.*, 2014; *Cael and Seekell*, 2016]. The contrast between abundance and surface area is due to the departure from a power law abundance-size relationship for small lakes. This departure occurs at scales where Earth's topography loses its fractal characteristics [*Cael and Seekell*, 2016]. The contrast in patterns between area and volume relates to the change in dimension. The connections between these geometric properties and macroscale patterns for lake ecosystems are unlikely to be uncovered by current approaches, which typically focus on chemical monitoring data [e.g., *Fergus et al.*, 2011; *Seekell et al.*, 2014]. Connecting geometric constraints influencing the size distributions of volume and area to geographic patterns is a new frontier in understanding the macroecology of lakes and an area with strong potential for the development of testable hypotheses.

To date, the global-scale characterization of Earth's lakes has been mostly a descriptive, data-driven endeavor, and few testable hypotheses have been developed [*Seekell et al.*, 2013]. Our study connects data with theory to develop and test hypotheses for lake volume-area scaling. We show that scaling relationships are mechanistic and general, and therefore ideal for global lake volume and depth estimates. Our results highlight the relative scarcity of lake waters in the hydrosphere and suggest a greater areal coverage of small, shallower, systems. These findings represent a fundamental contribution to advancing global-scale limnology.

Acknowledgments
This paper is based on research supported by the Knut and Alice Wallenberg Foundation, the National Science Foundation Graduate Research Fellowship Program under grant 2388357, and the National Science Foundation, award OCE-1315201. The authors declare no conflicts of interests. All data used for the study are available from the references and/or databases cited within the text. The global lake area census data were provided via email by C. Verpoorter. All analysis herein was performed in MATLAB 2014b, and all scripts are available online at http://cael.space.

References

Bastviken, D., J. J. Cole, M. L. Pace, and M. C. Van de Bogert (2008), Fates of methane from different lake habitats: Connecting whole-lake budgets and CH_4 emissions, *J. Geophys. Res. 112*, GO2024, doi:10.1029/2007JG000608.

Beauchamp, J. J., and J. S. Olson (1973), Corrections for bias in regression estimates after logarithmic transform, *Ecol*, *54*, 1403–1407.

Birkett, C. M., and I. M. Mason (1995), A new global lakes database for a remote sensing program studying climatically sensitive large lakes, *J. Great Lakes Res.*, *21*, 307–318.

Budyko, M. I (1980). *Climate: Present and Future* [in Russian], Gidrometeoizdat.

Cael, B. B., and D. A. Seekell (2016), The size-distribution of Earth's lakes, *Sci. Rep.*, *6*, 29,633.

Canfield, D. E., and R. W. Bachman (1981), Prediction of total phosphorus concentrations, chlorophyll *a*, and Sechi depths in natural and artificial lakes, *Can. J. Fish. Aquat. Sci.*, *38*, 414–423.

Carpenter, S. R. (1983), Lake geometry: Implications for production and sediment accretion rates, *J. Theor. Biol.*, *105*, 273–286.

Cox, B. L., and J. S. Y. Wang (1993), Fractal surfaces: Measurement and applications in the Earth sciences, *Fractals*, *1*, 87–115.

Dodds, P. S., and D. H. Rothman (2000), Scaling, universality, and geomorphology, *Annu. Rev. Earth Planet. Sci.*, *28*, 571–610.

Dodds, W. K., and M. R. Whiles (2010), *Freshwater Ecology*, Elsevier.

Downing, J. A., et al. (2006), The global abundance and size distribution of lakes, ponds, and impoundments, *Limnol. Oceanogr.*, *51*, 2388–2397.

Eilers, J. M., D. F. Brakke, D. H. Landers, and W. S. Overton (1989), Chemistry of lakes in designated wilderness areas in the western United States, *Environ. Monit. Assess.*, *12*, 3–21.

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.

Feng, M., J. O. Sexton, S. Channan, J. R. Townshend (2016), A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic-spectral classification algorithm, *Int. J. Digit. Earth*, *9*, 113–133.

Fergus, C. E., P. A. Soranno, K. S. Cheruvelil, and M. T. Bremigan (2011), Multiscale landscape and wetland drivers of lake total phosphorus and water color, *Limnol. Oceanogr.*, *56*, 2127–2146.

Goodchild, M. F. (1988), Lakes on fractal surfaces: A null hypothesis for lake-rich landscapes, *Math. Geol.*, *20*, 615–630.

Håkanson, L., and B. Karlsson (1984), On the relationship between regional geomorphology and lake morphometry: A Swedish example, *Geogr. Ann. A*, *66*, 103–119.

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, *Freshwater Biol.*, *52*, 814–822.

Heathcote, A. J., P. A. del Giorgio, and Y. T. Prairie (2015), Predicting bathymetric features of lakes from the topography of their surrounding landscape, *Can. J. Fish. Aquat. Sci.*, *72*, 643–650.

Herdendorf, C. E. (1982), Large lakes of the world, *J. Great Lakes Res.*, *8*, 379–412.

Hollister, J., and W. B. Milstead (2010), Using GIS to estimate lake volume from limited data, *Lake Reserv. Manage.*, *26*, 194–199.

Kanciruk, P., M. Gentry, R. McCord, L. Hook, J. Eilers, and M. D. Best (1987), National surface water survey: Western lake survey-phase I, data base dictionary, *ORNL/TM-10307*, Oak Ridge Natl. Lab.

Landers, D. H., W. S. Overton, R. A. Linthurst, and D. F. Brakke (1988), Eastern Lake Survey: Regional estimates of lake chemistry, *Environ. Sci. Technol.*, *22*, 128–135.

Lehner, B., and P. Döll (2004), Development and validation of a global database of lakes, reservoirs and wetlands, *J. Hydrol.*, *296*, 1–22.

Losev, K. S. (1989), *Water* [in Russian], Gidrometeoizdat.

L'Vovich, M. I. (1974), World Water Resources and Their Future [in Russian], Mysl.

Margalef, R. (1983), *Limnologia*, Ediciones Omega.

Mark, D. M., and P. B. Aronson (1984), Scale-dependent fractal dimensions of topographic surfaces: An empirical investigation, with applications in geomorphology and computer mapping, *Math. Geol.*, *16*, 671–683.

Martin, L. (1965), *The Physical Geography of Wisconsin*, Univ. of Wisconsin Press.

McClean, C. J., and I. S. Evans (2000), Apparent fractal dimensions from continental scale digital elevation models using variogram methods, *Trans. GIS*, *4*, 361–378.

McDonald, C. P., J. A. Rover, E. G. Stets, and R. G. Striegl (2012), The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent, *Limnol. Oceanogr.*, *57*, 596–606.

Meybeck, M. (1995), Global distribution of lakes, in *Physics and Chemistry of Lakes*, edited by D. Lerman, M. Imboden, and J. R. Gat, pp. 1–35, Springer.

Mulholland, P., and J. Elwood (1982), The role of the lake and reservoir sediments as sinks in the perturbed global carbon cycle, *Tellus*, *34*, 490–499.

Nace, R. L. (1969), In Water, Earth, and Man, edited by R. J. Chorley, p. 31–42, Methuen & Co., London.

Oliver, S. K., P. A. Soranno, C. E. Fergus, T. Wagner, L. A. Winslow, C. E. Scott, K. E. Webster, J. A. Downing, and E. H. Stanley (2016), Prediction of lake depth across a 17-state region in the United States, *Inland Waters*, *6*, 314–324.

O'Sullivan, P. E., and C. S. Reynolds (2005), *The Lakes Handbook*, Blackwell.

Ouchi, S., and M. Matsushita (1992), Measurement of self-affinity on surfaces as a trial application of fractal geometry to landform analysis, *Geomorphology*, *5*, 115–130.

Penck, A. (1894), *Morphologie der Erdoberfläche, vols I and II*, Engelhorn.

Raymond, P. A., et al. (2013), Global carbon dioxide emissions from inland waters, *Nature*, *503*, 355–359.

Renard, F., T. Candela, and E. Bouchaud (2013), Constant dimensionality of fault roughness from the scale of micro-fractures to the scale of continents, *Geophys. Res. Lett.*, *40*, 83–87, doi:10.1029/2012GL054143.

Ryanzhin, S. V. (2005), New estimates for global surface area and volume of natural world lakes, *Doklady Earth Sci.*, *401*, 253–257.

Ryanzhin, S. V. (2003), Reports of the 4th International Lake Ladoga Symposium (Joensuun Yliopisto, Joensuu, 2003), pp. 496–502.

Sachs, L. (1997), *Angewandte Statistik*, Springer.

Seekell, D. A., and M. L. Pace (2011), Does the Pareto distribution adequately describe the size-distribution of lakes?, *Limnol. Oceanogr.*, *56*, 350–356.

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.

Seekell, D. A., J. Lapierre, M. L. Pace, C. Gudasz, S. Sobek, and L. J. Tranvik (2014), Regional-scale variation of dissolved organic carbon concentrations in Swedish lakes, *Limnol. Oceanogr.*, *59*, 1612–1620.

Shiklomano, I. A. (1993), In Water In Crisis: A Guide to the World's Fresh Water Resources, p. 13–24, Oxford Univ. Press, New York.

Smith, R. J. (1993), Logarithmic transformation bias in allometry, *Am. J. Phys. Anthropol.*, *90*(2), 215–228.

Sobek, S., J. Nisell, and J. Fölster (2011), Predicting the volume and depth of lakes from map-derived parameters, *Inland Waters*, *1*, 177–184.

Tamrazyan, G. P. (1974), The total lake water resources of the planet, *Bull. Geol. Soc. Finl.*, *46*, 23.

Verpoorter, C., T. Kutser, and L. J. Tranvik (2012), Automated mapping of water bodies using Landsat multispectral data, *Limnol. Oceanogr. Methods*, *10*, 1037–1050.

Verpoorter, C., T. Kutser, D. A. Seekell, and L. J. Tranvik (2014), A global inventory of lakes based on high-resolution satellite imagery, *Geophys. Res. Lett.*, *41*, 6396–6402, doi:10.1002/2014GL060641.

Vollenweider, R. A. (1975), Input-output models with special reference to the phosphorus loading concept in limnology, *Schweiz. Z. Hydrol.*, *37*, 53–84.

Wagner, T., P. A. Soranno, K. S. Cheruvelil, W. H. Renwick, K. E. Webster, P. Vaux, and R. J. F. Abbitt (2008), Quantifying sample biases of inland lake sampling programs in relation to lake surface area and land use/cover, *Environ. Monit. Assess.*, *141*, 131–147.

Wetzel, R. G., and G. E. Likens (2000), *Limnological Analyses*, Springer, New York.

Wetzel, R. G. (2001), *Limnology: Lake and River Ecosystems*, Academic Press.