The distribution of depth, volume, and basin shape for lakes in the conterminous United States

Keenan J. Ganz[1,2], Max R. Glines[1], and Kevin C. Rose[1,*]

[1]Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY, 12180

[2]School of Environmental and Forest Sciences, University of Washington, Seattle, WA, 98195

[*]Corresponding author; rosek4@rpi.edu

Abstract

Depth regulates many attributes of aquatic ecosystems, but relatively few lakes are measured, and existing datasets are biased toward large lakes. To address this, we used a large dataset of maximum (Z_{max}; n = 16,831) and mean (Z_{mean}; n = 5,881) depth observations to create new depth models, focusing on lakes < 1,000 ha. We then used the models to characterize patterns in lake basin shape and volume. We included terrain metrics, water temperature and reflectance, polygon attributes, and other predictors in a random forest model. Our final models generally outperformed existing models (Z_{max} R^2 = 0.35; RMSE = 8.0 m and Z_{mean} R^2 = 0.36; RMSE = 3.0 m). Our models show that lake depth followed a Pareto distribution, with 2.8 orders of magnitude fewer lakes for an order of magnitude increase in depth. Additionally, despite orders of magnitude variation in surface area, most size classes had a modal maximum depth of ~5 m. Concave (bowl-shaped) lake basins represented 79% of all lakes, but lakes were more convex (funnel-shaped) as surface area increased. Across the conterminous US, 9.8% of all lake water

24 was within the top meter of the water column, and 48% in the top 10 m. Excluding the
25 Laurentian Great Lakes, we estimate the total volume in the conterminous US is 1,057 to 1,294
26 km^3, depending on whether Z_{max} or Z_{mean} was modeled. Lake volume also exhibited substantial
27 geographic variation, with high volumes in the upper Midwest, Northeast, and Florida and low
28 volumes in the southwestern US.

1 Introduction

Depth is an essential attribute of aquatic ecosystems. For example, many physical characteristics covary with depth. Deeper lakes often have larger heat budgets (Birge 1915) and therefore take longer to freeze in autumn (Walsh et al. 1998; Kirillin et al. 2012). Indeed, the timing of seasonal ice cover onset is closely related to depth (Efremova and Pal'shin 2011). Lake depth also affects mixing regimes through its impact on the stability of the water column (Håkanson 2004). Deep lakes may remain stratified for longer periods than shallow lakes, concentrating solar radiation in the uppermost water layers (Calamita et al. 2021).

In addition to regulating lake physics, depth modulates numerous chemical and biological features of lakes. In a meta-analysis of lakes receiving agricultural runoff, deeper lakes had lower surface phosphorus than shallow lakes (Taranu and Gregory-Eaves 2008). A later analysis of 1,026 randomly sampled US lakes identified maximum depth as the most important predictor of total nitrogen, total phosphorus, turbidity, and dissolved organic carbon concentrations (Read et al. 2015). These water quality metrics are known to influence numerous other aspects of lakes, including productivity and methane emissions (West et al. 2016).

Despite the importance of lake depth, direct measurements are relatively uncommon, and few lakes have well-resolved bathymetric maps. When measurements exist, they often include only a maximum depth, preventing widespread assessments of the hypsography of lake basins. Additionally, existing databases exhibit bias towards lakes that are large, well-studied, and easy to sample (Hanson et al. 2007; Stanley et al. 2019). In contrast with surface area, lake depth cannot be directly measured from remote sensing products yet is just as important for

characterizing lake volume. Therefore, there is a critical need for general-use predictive models of lake depth, especially for smaller lakes.

Existing models of maximum lake depth (Z_{max}) have generally taken one of two conceptual approaches by focusing on either fractal patterns or empirical relationships to predict lake depth. In the first, the fractal nature of Earth's topography is used to develop scaling relationships between lake surface area and basin morphometry. The fractal approach to lake depth modeling assumes that Earth's topography is an idealized self-affine surface. On such surfaces, topographic scaling is related to horizontal scaling via the Hurst exponent (H) (Dodds and Rothman 2000). These assumptions result in scaling rules that relate lake area with volume ($V \sim A^{1+\frac{H}{2}}$) and depth ($Z \sim A^{\frac{H}{2}}$) (Carpenter 1983; Wetzel and Likens 2000). Therefore, the Hurst exponent may be estimated by doubling the slope of a log-log regression between lake depth and lake area. Recent empirical results from large lake databases have good agreement that $H = 0.4 \pm 0.1$ (Cael et al. 2017; Cael and Seekell 2022). However, it must be emphasized that these scaling rules apply to collections of lakes and do not constitute predictive models for individual lakes (Cael and Seekell 2022).

In the second type of model, empirical relationships exploit correlations between lake and watershed attributes to predict depth. Empirical models of Z_{max} or mean depth (Z_{mean}) have generally relied upon correlations with lake polygon attributes, especially lake surface area, and regional terrain characteristics (e.g., local slope) under the assumption that nearshore topography is formed by the same processes which formed the lake basin (Hutchinson 1957). For example, a

74 geometric approximation of Z_{max} may be derived from nearshore slope and distance to lake
75 center with a simple cone model (Equation 1; S: nearshore slope; D: distance to lake center).

$$Z_{max} = \tan S \times D \quad (1)$$

76 The cone model may also be applied to estimate lake volume when lake surface area is known
77 (Equation 2; A: lake area). Under this idealized geometry, Z_{mean} is equal to one third of Z_{max} or,
78 equivalently, to the ratio of lake volume to lake area.

$$V = \frac{A \times Z_{max}}{3} \quad (2)$$

79 Alternatively, a box model may be used if mean depth is known. In this case, lake volume is the
80 product of lake area and mean depth. The ratio between Z_{max} and Z_{mean} can provide information
81 on the relative shape of a lake's basin, for example, whether it is concave (i.e., bowl-shaped) or
82 convex. Thus, modeling both Z_{max} and Z_{mean} simultaneously may provide critical additional
83 information to characterize the degree to which a simple cone model represents reality.
84
85 Lake and watershed attributes that can be remotely sensed may provide the capability to improve
86 empirical lake depth models beyond simple cone model estimates alone. For example, estimates
87 of lake volume, water temperature, and water clarity may provide additional information on lake
88 depth (Becker and Daw 2005; Khazaei et al. 2022). But it is not currently clear which, if any, of
89 these predictor categories is most important. One of the most accurate lake depth models to date
90 included estimated lake volume as a predictor of lake depth (Khazaei et al. 2022). However, this
91 study focused primarily on very large (mean 894 km^2, median 2.1 km^2) and deep (mean 34.29 m,
92 median 13.10 m) lakes, where measured lake depth is often known. Nevertheless, Khazaei et al.

93 (2022) demonstrated how estimated parameters can improve lake depth predictions and we
94 include a depth estimate as a predictor in our models.
95
96 The compilation and release of large numbers of observed lake depths permits the creation and
97 validation of novel lake depth models. For example, the release of LAGOS-US DEPTH
98 (Stachelek et al. 2021; Webster et al. 2022), a dataset with >15,000 observations of Z_{max} across
99 the United States marks a prime opportunity to create generalized lake depth models over a large
100 geographic region (Figure 1). Here, we leveraged machine learning approaches and this large
101 publicly available database of Z_{max} and Z_{mean} to generate models of lake depth at the
102 conterminous US scale and compare our model to a suite of existing lake depth models. We use
103 our modeled lake depths, which were more accurate than prior model predictions, to calculate a
104 Hurst exponent, identify which variables are most predictive of lake depth, and, by calculating
105 the ratio between Z_{max} and Z_{mean}, the frequency of convex versus concave lake bathymetries
106 across the US. We then use our model to estimate the volume of all lakes across the US and map
107 their geographic variations.

108 ## 2 Methods

109 ### 2.1 Overview

110 Our approach to modeling and quantifying large-scale patterns in lake maximum and mean depth
111 proceeded in a multi-step process. First, we obtained lake polygons and depth measurements
112 from the publicly available Lake Multi-Scaled Geospatial and Temporal Database (LAGOS-US;
113 Cheruvelil et al., 2021; Stachelek et al., 2021). This data set contained *in situ* measurements of
114 maximum lake depth for 17,675 lakes, mean lake depth for 6,137 lakes, and 479,950 lake

115 polygons at least 1 ha in area across the conterminous US. LAGOS-US excluded the Laurentian
116 Great Lakes. We implemented two modeling approaches. First, we excluded lakes larger than
117 1,000 ha (n = 1,128 of all polygons) because the majority of depths of lakes > 1,000 ha are
118 already known (73% for our lake data set; Supplemental Information Figure 1) and large lakes
119 are typically outliers with respect to depth (Sobek et al. 2011; Oliver et al. 2016). We also
120 excluded extremely shallow lakes (< 0.5 m maximum depth, n = 8), resulting in a total maximum
121 and mean depth data set of 16,831 lakes and 5,881 lakes, respectively. In a second modeling
122 approach, we included all lakes > 1,000 ha (excluding the Laurentian Great Lakes) to quantify
123 the total volume of lakes in the conterminous US (see Supplemental Information). These
124 modeling approaches were motivated by our hypothesis that small lakes are formed by a larger
125 variety of processes, while large lakes are generally formed by fewer and similar processes that
126 form the surrounding topography. Therefore, we hypothesized that a simple cone model may be
127 sufficient to estimate the depth of large lakes, but more covariates were necessary to accurately
128 estimate the depth of small lakes.
129
130 We next compiled a suite of 165 predictor variables for all lakes that we hypothesized may be
131 predictive of lake depth. All predictor variables were attributes that could be obtained from
132 continental-scale data products (e.g., digital elevation models and remote sensing). Missing
133 values were filled by median imputation for at most ~2% of observations. We organized
134 predictors into seven categories, including (1) lake polygon attributes, (2) local terrain metrics,
135 (3) neighborhood terrain metrics, (4) regional (HUC4) summary statistics, (5) water surface
136 temperature and (6) optical properties, and (7) a suite of other variables from LAGOS-US (Table
137 1). Each predictor variable category is described in the following sections and a complete list of

all predictor variables is provided in Supplemental Information. All predictors were identical in the maximum and mean depth models, with the exception of those derived from thermal and optical satellite imagery (see below).

Once we compiled all predictor variables, we implemented a random forest model (Breiman 2001; Wright and Ziegler 2017) to predict Z_{max} and Z_{mean} from the predictor variables. We next implemented a pruning procedure to identify the most parsimonious model. Once a final model was selected, we assessed model accuracy relative to previous models (Table 2) and characterized large-scale patterns in lake depth and volume. Comparison with previous depth models often necessitated some modification to previous models, which are described below. We performed spatial analyses in QGIS 3.22 (QGIS Development Team 2021), calculated remote sensing metrics in Google Earth Engine (GEE; Gorelick et al., 2017), and performed model fitting in R 4.1.2 (R Core Team 2021) with the mlr3 framework (Lang et al. 2019).

2.2 Lake polygon attributes

We included five lake polygon attributes as predictor variables. Lake size and shape have previously been shown to be predictive of lake depth (Håkanson and Karlsson 1984) but the specific shape attributes most predictive of lake depth are unclear. The attributes we selected

included lake area, perimeter, and a composite of these values, the shoreline development index (Equation 3; P: lake perimeter).

$$SDI = \frac{P}{2\sqrt{\pi A}} \quad (3)$$

We also included the length to width ratio of the minimum oriented bounding box and the maximum distance from shoreline (Garcia-Castellanos and Lombardo 2007). We calculated all polygon attributes in QGIS using an Albers equal area projection (EPSG 5070).

2.3 Local terrain predictors

We included 49 local terrain attributes as predictor variables. We included a range of terrain attributes because nearshore watershed characteristics have been shown to be closely associated with lake depth (e.g. Håkanson 2004), but the attributes that are most predictive are unclear. In this approach, terrain attributes are calculated for each pixel of a digital elevation model within some buffer zone around a lake but excluding the lake surface itself. This distribution of values is then summarized by its median, minimum, maximum, and standard deviation. Finally, the summary value was used to predict lake depth. A complete list of all predictor variables used in this study is provided in an accompanying data package.

We included two broad categories of terrain attributes: local and neighborhood. These categories differed in how the terrain attributes were calculated for each pixel in the buffer region. Our local terrain attributes were derived from the eight pixels surrounding each pixel in the buffer region, while our neighborhood terrain attributes were derived from circular kernels ranging from 50 m to 1,000 m in radius. In both categories, we computed terrain attributes for each pixel in the buffer region, then summarized the distribution of attribute values to derive specific predictor

variables. The number of predictors in this category should not be taken as an exhaustive inventory of all possible terrain attributes. Given the combinatoric nature of terrain attributes, kernel size, and summarization functions, many other predictor variables may be possible.

We calculated both local and neighborhood terrain analyses in the GEE environment using the Forest and Buildings Removed Copernicus 30 m DEM (FABDEM; Hawker et al., 2022). We calculated terrain attributes within 100 m buffer zones around each lake polygon. Predictor variables included the minimum, maximum, median, and standard deviation for each attribute within the buffer. Because GEE sets the maximum polygon shape complexity at 200,000 edges, we occasionally had to reduce the number of polygon edges using an iterative endpoint fit algorithm with a 30 m tolerance (Douglas and Peucker 1973).

We considered classic terrain attributes, such as slope and terrain roughness, as well as a variety of curvature metrics using the Terrain Analysis in GEE package (Safanelli et al. 2020). Vertical curvature may be thought of as the second derivative of elevation in the direction of maximum slope. We hypothesized that curvature may correct for error in the cone model due to basin shape (Equation 4; C: nearshore vertical curvature). As defined in Equation 1 above, here S is the nearshore slope and D is the distance to lake center.

$$Z_{max} = \tan S \times D + \frac{C}{2} \times D^2 \quad (4)$$

We included both the linear and quadratic terms in Equation 4 as separate predictors in the local terrain metrics category. In addition to these local terrain features, we derived estimates of mean lake depth and volume following Model 5 in Messager et al. (2016). This model is a simple

197 linear combination of nearshore slope and lake area. We hereafter refer to these predictors as the
198 geostatistical depth estimate and the geostatistical volume estimate.

199 **2.4 Neighborhood terrain predictors**

200 We included 30 neighborhood terrain attributes as predictor variables. Local terrain metrics
201 described topographic features in the 8-neighborhood of a particular pixel. To include additional
202 topographic information, we constructed circular kernels around each pixel in the lake buffer
203 region and compared the elevation value at the kernel center with all other elevation values
204 within the kernel. Because it is not known what neighborhood spatial extent best predicts lake
205 depth, we repeated this procedure for kernels of radius 50 m, 100 m, 500 m, and 1,000 m to
206 capture topographic features at these spatial extents (M. Halabisky, unpubl.). We then calculated
207 two neighborhood metrics: terrain deviation and terrain relief. Terrain deviation is the Z-score of
208 the kernel center in the distribution of pixels included in the kernel. Terrain relief is the
209 maximum absolute difference between the kernel center pixel and all other pixels in the kernel.
210 Both metrics describe the relative relief of terrain, with negative values indicating valleys and
211 positive values indicating ridges (Supplemental Information Figure 2).

212 **2.5 HUC4 summary statistics**

213 We included 40 summary statistic attributes as predictor variables. Consistent with past research
214 (Oliver et al. 2016), we hypothesized that lakes would exhibit spatial autocorrelation within the
215 same 4-digit hydrologic unit code (HUC4) unit in the US Geological Survey's National
216 Hydrography Dataset. We grouped lake polygon attributes, buffer terrain attributes, maximum
217 depth, and relative depth (Håkanson, 1981; reproduced as Equation 5 below; RD: relative depth)

by HUC4 unit and summarized these attributes by minimum, maximum, median, and standard deviation. Each summary statistic was included as a separate predictor.

$$RD = \frac{Z_{max}\sqrt{\pi}}{20\sqrt{A}} \quad (5)$$

2.6 Water surface temperature predictors

We included 12 water surface temperature attributes as predictor variables. Previous research indicates that lake surface water temperatures are related to lake depth and volume. For example, deeper lakes may be relatively warm during seasons of heat loss to the atmosphere compared with shallower lakes because it takes longer for them to cool (Becker and Daw 2005). However, the best month or months to calculate this temperature anomaly is currently unclear. Therefore, we estimated temperature anomalies for every lake relative to regional air temperature on a monthly basis. To do this, we first estimated water surface temperature from the thermal bands 10 and 11 in cloud-masked Landsat 8 imagery (Roy et al. 2014). We converted top-of-atmosphere radiances to brightness temperature and atmospherically corrected the data according to the split window algorithm in Du et al. (2015) for the widest range of column water vapor values and band-specific water emissivity values in Vanhellemont (2020). We then calculated monthly mean water surface temperature at the point furthest from shore (for maximum depth) or all lake pixels (for mean depth) in each lake polygon across all available years (2013 – present; $\bar{n} \approx 15$ per lake per month; Supplemental Information Figure 3). We next estimated the monthly mean air temperature 2 m above ground using ERA5 atmospheric reanalysis data from the Copernicus Climate Data Store (Muñoz Sabater 2018). We selected this atmospheric dataset since its coarse resolution (30 km) would be resistant to influence by large lakes. Finally, we

calculated the monthly lake temperature anomaly as the difference between the monthly mean air temperature and monthly mean water temperature.

2.7 Water surface reflectance

We included 10 water surface reflectance attributes as lake depth predictor variables. Surface reflectance is associated with variations in the concentration of optically active substances, including dissolved organic matter, suspended sediments, and chlorophyll. Because the concentrations of optically active substances scales with lake size (Read et al. 2015), we hypothesized that water surface reflectance attributes may be predictive of lake depth. Additionally, optical estimates have previously been used to estimate water depth in marine environments using the ratio of log-transformed surface reflectance in different wavelengths (Stumpf et al. 2003; Geyman and Maloof 2019). To obtain these attributes, we calculated the mean reflectance for each Landsat 8 band for all images captured May through July of each year (to avoid ice cover) at the pixel furthest from shore in each lake polygon ($\bar{n} \approx 43$ per lake; Supplemental Information Figure 3). We log-transformed these reflectance values and calculated all unique pairwise band ratios as separate predictor variables, consistent with reflectance ratios used in coastal bathymetry mapping (Geyman and Maloof 2019).

2.8 Modeling Workflow

After compiling all predictor variables, we performed feature selection to ensure a parsimonious final model. We did this in three steps. First, we removed candidate predictor variables with poor predictive power from the random forest model using the Boruta algorithm in the Boruta R package (Kursa and Rudnicki 2010). The Boruta algorithm proceeds by comparing the importance of a variable in a random forest model with random noise. We retained predictors that significantly outperformed ($p < 0.01$) random noise while all others were excluded. This

process removed 41 predictor variables from the maximum depth model and 84 predictors from the mean depth model. In our second step, we aggregated median variable importance metrics from the first step and excluded variables in the bottom 80% of permutation importance from the Boruta run (99 excluded from maximum depth, 64 excluded from mean depth). In our third step, we eliminated any predictors which had a Pearson correlation coefficient greater than 0.7 with any other predictor retained after the second step (8 excluded from maximum depth, 4 excluded from mean depth). Although random forest models are resistant to multicollinearity, we reasoned that including well-correlated predictors increases model complexity with little increase in predictive power. Therefore, the final set of predictors in our model should be interpreted as one of several possible combinations that would produce similar predictive power.

Since the models we identified from the literature were designed to predict maximum depth, it was not appropriate to compare our mean depth model with these. Instead, we report model performance statistics for our mean depth model, the random forest model from Khazaei et al. (2022) trained on mean depth data, and the mean depth estimate from Messager et al. (2016).

We used the final random forest models to predict Z_{max} and Z_{mean} for all lakes smaller than 1,000 ha in LAGOS-US LOCUS without Z_{max} or Z_{mean} observations. We report predicted maximum depths, mean depths, lake volumes as calculated from Equation 2, volume development ratios as calculated from Equation 3, Hurst exponents of modeled and measured depths derived from bootstrapped log-log regressions, variable importance metrics from Boruta runs, water volume associated with sections of the water column, and fits of the Pareto distribution to counts of lake depth. The coefficient describing the fit of a distribution to the Pareto distribution can provide

information on how concentrated a feature is in a population, with a more negative coefficient describing features that are more highly concentrated than those with a lower (less negative) coefficient. We truncated these fits by visual inspection to account for the minimum lake area mapped in the LAGOS dataset and relatively few observations of the largest lakes. To calculate the volume of a section of the water column, we assumed that all lakes had conic basins. This simplification yields an estimate of water volume for a section defined by depths z_1 and z_2 when lake area and maximum depth are known (Equation 6). The derivation of Equation 6 is provided in Supplemental Information.

$$V_{section} = \frac{A}{3Z_{max}^{\ 2}}[(Z_{max} - z_1)^3 - (Z_{max} - z_2)^3];\ z_1 < z_2 < Z_{max} \quad (6)$$

We assessed the utility of our model by comparing it to previous publications (Table 2). We implemented previously published models as closely as possible to original formulations, albeit with several modifications to conform to our lake data set. First, the temperature-based model described in Becker and Daw (2005) could not be replicated due to a lack of cloud-free nighttime land surface temperature imagery in the study area. Likewise, the bathymetry-based ("true slope") model of Stachelek et al. (2022) could not be replicated due to insufficient bathymetric data for our lake depth data set. Exact replication of prior work was further precluded by differing elevation data sources, processing constraints, and a lack of consistent watershed boundaries. The major difference between the present study and previous publications was the use of a constant 100 m buffer to calculate terrain metrics. For example, Heathcote et al. (2015) used a variable buffer distance proportional to the square root of lake area while Sobek et al. (2011) used constant buffers ranging from 50 m to 1,000 m. Our constant 100 m buffer represents roughly half of the mean distance from lake edge to center in our lake data set and balances the need for sufficient DEM pixels while maintaining emphasis on nearshore

topography. To test if the size of the buffer zone was an important factor controlling model accuracy, we constructed separate models with variable buffer sizes. Overall model predictive power did not change with buffer size nor when a variable buffer size was used instead of a constant buffer size (data not shown). Our experiments with other buffer sizes were consistent with the results in Zhan et al. (2023), which found no difference in the performance in a variety of lake depth models trained on the LAGOS dataset. We hypothesize that buffer size controls model performance in regional studies, but at continental scale this choice is less important.

All models except that of Oliver et al. (2016) were evaluated in a spatial cross-validation design to prevent biased performance metrics. Spatial cross-validation is similar to naïve cross-validation but uses k-means clustering to ensure that the resulting resamples are spatially disjoint (Brenning 2012). This prevents spatial autocorrelation from introducing bias in model performance metrics (Lovelace et al. 2022). We constructed our training procedure such that each model was trained 64 times total. Models that did not require hyperparameter tuning were evaluated with eight spatial cross-validation clusters eight times (8 x 8 = 64). Models that did require hyperparameter tuning were evaluated with four spatial cross-validation clusters four times. Within each spatial cross-validation cluster, hyperparameters were tuned were tuned over four iterations (4 x 4 x 4 = 64). We made one exception to this procedure with the linear mixed effects model of Oliver et al. (2016). This model explicitly incorporated spatial autocorrelation via a geographic blocking design. Therefore, stratified cross-validation was used over spatial cross-validation because spatially disjoint resamples are not a true measure of this particular model's performance. We also acknowledge that controversy remains as to which model

evaluation procedure is best for machine learning models of ecological variables (Meyer and Pebesma 2022).

To assess the accuracy of our model and previous models we calculated a suite of performance metrics during the aforementioned cross-validation procedures. Evaluation metrics included R^2, root mean square error (RMSE), mean absolute error (MAE) and percentage bias (PBIAS). Each model was trained 64 times total in our design. From each fit, we computed model performance metrics and we report the mean model performance across all fits. Final model predictions were derived from models fit on all data with the optimal hyperparameter values found during spatial cross-validation. We calculated residuals for the final models and report them for the entire conterminous US and by Level 1 ecoregions, as defined by the US EPA (Omernik and Griffith 2014).

Once we had functional Z_{max} and Z_{mean} depth models, we calculated the ratio between Z_{max} and Z_{mean} to assess the basin shape of each lake. Håkanson (2004) quantified deviations from the cone model via the volume development ratio, also called the form factor (Equation 7). Using the concavity definition used for hypsometric integrals provided in Håkanson (1977), convex lakes

have $V_d < 1.5$ while concave lakes have $V_d > 1.5$. Under this classification schema, it is noted that a perfectly conical lake, i.e. $V_d = 1$, is in fact classified as moderately convex.

$$V_d = 3\frac{Z_{mean}}{Z_{max}} \quad (7)$$

3 Results

After pruning, our maximum depth random forest model contained 18 predictor variables (Table 3). This is more predictors than contained in earlier models but resulted in the highest R^2 (0.35) second-lowest MAE (5.02 m), and a percent bias and RMSE in line with past models (-64% and 8.04 m; Table 3). The mean depth random forest model performed similarly well, but with 13 predictor variables (Table 3). In our model, as well as our implementation of past lake depth models, depth prediction accuracy decreased as depth increased (Figure 2). Small lake depths were slightly overestimated, while large lake depths were underestimated (Supplemental Information Figure 4). Hyperparameter tuning in our maximum depth random forest model identified optimal hyperparameters as a sample fraction of 0.28, 13 variables tried per split, and a minimum node size of 3 observations. Our final model showed that residuals were of similar magnitude across all ecoregions with the exception of Mediterranean California (Supplemental Information Figure 5 and 6).

Bootstrapped log-log regressions of measured maximum depths vs. lake area from LAGOS-US DEPTH indicate the Hurst exponent is 0.395 with a 95% confidence interval of (0.382, 0.408). However, when we repeated our analysis with pooled modeled and observed depths, we found a Hurst exponent of 0.160 (0.158, 0.163), indicating a substantial difference between the coefficient calculated from modeled versus observed depths. To determine whether the

366 difference was caused by large numbers of small lakes (that have not been included in previous
367 calculations of the Hurst coefficient) or the random forest model, we ran a third bootstrap on
368 modeled lake depths using only lakes that had a similar area distribution to observed lake depths.
369 In this case, the Hurst exponent was 0.126 (0.124, 0.129). Predictions from other models showed
370 similar deviations from the empirical Hurst exponent, with the exception of the mean depth
371 model from Messager et al. (2016) (Table 3).
372
373 The coefficient describing the fit of lake depth frequencies to the Pareto distribution was more
374 negative than it was for lake area frequencies (-2.79 versus -0.95). Similarly, the coefficient
375 describing lake area frequencies was more negative than it was for lake volume frequencies (-
376 0.95 versus -0.76). For comparison, Downing et al. (2006) reported a coefficient of -1.06
377 describing the distribution of lake area frequencies, which is close to our estimate of -0.95.
378
379 Our results show that at least one predictor variable from all categories was retained in the final
380 model (Table 1; Figure 3). Both satellite-derived predictor categories were well-represented in
381 the final model, with 50% of monthly surface temperature measurements and 20% of surface
382 water reflectance retained in the final model. Temperature data from the shoulder seasons were
383 most likely to be retained while the B2:B4 and B3:B5 reflectance ratios were retained.
384 Neighborhood terrain attributes were the most likely to be removed by our pruning procedure
385 with only one predictor (median 50 m terrain deviation) out of 32 retained in the final model.
386 Based on permutation importance, water surface temperature predictor variables during shoulder

387 seasons were the most important to the overall model (Figure 3; see also Supplemental
388 Information).
389
390 Maximum depths for modeled lakes smaller than 1,000 ha and all lakes with depth observations
391 (n = 477,661) followed a Gaussian distribution, with increasing skew as lake area increased
392 (Figure 4). All lake area classes had a mode depth of approximately 5 m. In aggregate, modeled
393 and known lake depths closely followed a Pareto distribution except at the very shallowest and
394 deepest lakes (Figure 5). Our mean depth model predicted on average 97% greater volume than
395 the maximum depth model, and bootstrapped regressions between the two volume predictions
396 had a median slope of 0.77 (Supplemental Information Figure 7). By dividing total lake volume
397 by total lake area, we calculated that the mean lake among our modeled and observed lakes is
398 9.07 m deep. Visualization of lake volume across the US shows substantial variation at broad
399 extents, with the largest lake volumes in areas such as in the upper Midwest, Northeast, and
400 Florida (Figure 6). Among lakes, our modeling results indicate that 48% all lake water is within
401 the shallowest 10 m of water depth, and 9.8% is in the top meter alone, owing to the vastly larger
402 number of shallow lakes than deep lakes (Figure 7).
403
404 By combining our maximum and mean depth models, we found that the distribution of the
405 volume development ratio was dependent on lake area (Figure 8). Our results show that concave
406 lakes are far more numerous than convex lakes. We estimate that 378,179 (79%) lakes have a
407 volume development ratio indicative of a concave shape, while 98,910 (21%) are convex
408 (median volume development ratio of 1.9). However, basin shape varied with lake size. Lakes
409 with area between 10^4 and 10^5 m^2 were much more likely to be highly concave, with a mode

410 volume development ratio of approximately 2. Overall, the volume development ratio decreased
411 (i.e., lakes become more funnel-shaped) as lake area increased.
412
413 We used separate models trained on all available depth data, including lakes > 1,000 ha, to
414 estimate the total volume of all lakes in the conterminous US (see Supplemental
415 Information for further details). Using this maximum depth model, we estimate there is a total
416 lake volume of 1057 km^3 in the conterminous US. Using this mean depth model, we estimate
417 there is a total lake volume of 1294 km^3. Predictions from these and other models (n = 9; see
418 Table 3) also trained on the same lakes predicted volumes ranging from 286 km^3 to 1,616 km^3,
419 with a mean of 991 km^3 (standard error: 134 km^3; Supplemental Table 1). The largest lakes by
420 area had the most influence on these estimations. For example, the 100 largest lakes by area
421 account for over 50% of the total lake volume in the conterminous US.

422 **4 Discussion**

423 Our lake depth models leverage large remotely sensed data and an *in situ* database of depth
424 measurements to create, by several measures, the most accurate lake depth models to date (Table
425 3). These improvements are derived from an increase in the overall number of predictors as well
426 as the inclusion of remote sensing data products as yet unused to model lake depth. However, it
427 is important to note that models are not necessarily improved just because of a larger number of
428 potential predictor variables. In our study, we integrated a diverse suite of predictor types,
429 including local and neighborhood terrain shape, lake shape, temperature, optical, and regional
430 characteristics that we hypothesized were related to lake depth. The overall improvements in lake

depth characterization, including for both maximum and mean depth, permit new insights into this essential lake attribute across the conterminous United States.

While our model represents an improvement over existing lake depth models, it is far from perfect, and there are inherent limitations that prevent accurate depth characterization in all instances. For example, Stachelek et al. (2022) used bathymetric models to show that inaccurate nearshore slope measurements introduce error in cone models of Z_{max} and concluded that even with complete bathymetric data, a cone model of maximum lake depth could only explain roughly 70% of the variance in Z_{max}. Other sources of uncertainty include factors such as the coarse cell size of FABDEM (30 m), changes in optical properties of lakes over time, and inconsistent measurement protocols among LAGOS-US DEPTH data sources.

Because the temperature anomaly (i.e., the difference between water and air temperature) was a top predictor of both Z_{max} and Z_{mean} (Figure 3), efforts to further improve lake depth predictions may benefit from refining temperature predictions (for both water and air over the lake). For example, our models may be improved by implementing controls for ice or modeling overall lake energy balance from gridded meteorological data. More generally, developing more accurate predictors, rather than increasing the number of predictors, is potentially a more promising direction for further improving lake depth models. It was also notable that another top predictor of lake depth was the Landsat B2:B4 ratio (Figure 3). This ratio has traditionally been used to characterize lake clarity (Lathrop 1992; Olmanson et al. 2016). It is unclear if the ratio is important because of bottom effects or because there is a relationship between lake size (surface

453 area or depth) and the concentration of optically active substances (e.g., algal biomass,
454 chromophoric dissolved organic matter).
455
456 We also note the large differences in reported model performance metrics (Table 2) and
457 measured model performance metrics (Table 3). These differences demonstrate the importance
458 of comparing lake depth models on a common dataset and the need to include spatial cross-
459 validation during model fitting (Lovelace et al. 2022). Without spatial cross-validation,
460 information leaks from the training to the validation data through spatial autocorrelation. This
461 can bias model performance metrics upward, as observed here.
462
463 Our volume estimates for all lakes, including those > 1,000 ha (but excluding the Laurentian
464 Great Lakes) yielded a total volume of 1,057 to 1,294 km^3 (using the maximum and mean depth
465 models, respectively). For comparison, this total volume of all lakes equates to about 10% of the
466 total volume of Lake Superior, the largest lake in the conterminous US (Matheson and Munawar
467 1978) highlighting just how unevenly depth and volume are distributed among lakes. These total
468 volume estimates were driven by massive volumes in a small number of large lakes. Although
469 depth data are unavailable for many small lakes, it is in fact the largest lakes that contain most of
470 the freshwater in the US. Additionally, we note that when including all lakes in our modeling
471 framework, including those > 1,000 ha, other past modeling approaches approach or even exceed
472 the performance of our model (Supplemental Information Table 1). For example, the model of
473 Hollister et al. (2011) has a lower root mean squared error (10.2 versus 12.0) and similar mean
474 absolute error (5.87 versus 5.90) and R^2 (0.32 versus 0.38) while also being more parsimonious.
475 The comparative advantage of our modeling approach is in predicting the depth and volume of

476 small lakes, which are much more numerous and less frequently measured than large ones. But
477 since total lake volume is driven by the largest lakes, future studies must note this distinction
478 when selecting a modeling approach. Overall, our results indicate that small lakes require greater
479 depth model complexity, while more parsimonious models derived from surrounding terrain
480 alone may be sufficient to model the depth of large lakes.
481
482 A comparable estimate for the total volume of lakes may be derived from values given in Table 2
483 in Cael et al. (2017). Assuming that each lake in our data set had the average volume for its size
484 class in the cited table, we calculate a total lake volume of 1,351 km^3 (details in Supplemental
485 Information). The difference in volume calculations derived from our model and estimated from
486 Cael et al. (2017) is associated with the fact that we report a lower Hurst exponent. A lower
487 Hurst exponent attributes larger depths to small lakes and smaller depths to large lakes than
488 would be otherwise predicted by a higher Hurst exponent. Ours and other models implemented
489 here overestimate shallow depths and underestimate deep depths, driving a decrease in the Hurst
490 exponent (Supplemental Information Figure 4).
491
492 Bootstrapped log-log regression of our modeled depths indicate a Hurst exponent much lower
493 than would be expected from empirical studies (Cael et al. 2017; Cael and Seekell 2022). Given
494 the remarkable consistency of fractal scaling relationships in nature, one would expect an
495 empirically derived model of lake depth to preserve this relationship. However, the fractal
496 scaling of Earth's surface is a statistical property of large collections of lakes and should not be
497 used to predict individual lake properties. That our random forest model and other lake depth
498 models in the literature do not preserve $H = 0.4$ is indicative that covariates other than lake area

499 are as or potentially more effective at predicting individual lake depths. Indeed, several other
500 predictor categories were of similar or greater importance compared with lake area in our models
501 (Figure 3).
502
503 The relationships between frequency and lake areas, depths, or volumes closely follow the Pareto
504 distribution, except at the tails of the distributions (Figure 5). Some of the deviation from the line
505 could be due to the fact that our lake model was focused on small lakes, which are traditionally
506 under-represented in observational studies. Because of this, the data included in Figure 5 for very
507 large lakes includes only those with observed measurements. While a large majority of very
508 large lakes have observed depths (Supplemental Information Figure 1), the lack of inclusion of
509 some very large lakes could contribute to the deviation away from the regression line. However,
510 deviations like this have been observed previously, including for very small lakes, too (e.g.
511 Seekell and Pace 2011). The fact that lake depth frequencies had the most negative coefficient
512 among our three regressions in Figure 5 means that lake depth was the feature that was most
513 concentrated in the fewest number of lakes. That is, most lakes are fairly shallow, whereas very
514 few are very deep. The small number of deep lake habitats has important biological and
515 ecological implications. For example, as climate change continues to warm lakes (Jane et al.
516 2021), deep habitats may become increasingly important as organisms seek to maintain their
517 preferred temperatures by either moving deeper or dispersing to a new area (Pinsky et al. 2013).

518 The fact that large depths are highly concentrated in few lakes means there may be few
519 opportunities for organisms to exploit deep habitats to avoid warming temperatures.
520
521 Our and previous models indicate that the relationship between lake area and lake depth is not as
522 strong as empirical studies suggest. For example, it is interesting to note that the modal depth for
523 all size classes is ~5 m, despite orders of magnitude variation in lake area (Figure 4). The fact
524 that lake depth and lake area are relatively uncoupled has many implications for the structure and
525 functioning of lake ecosystems, including for lake mixing regimes, light availability, and
526 dissolved oxygen. Because lake depth and lake area are not tightly coupled, habitats may be far
527 more heterogeneous across lakes than previously recognized. For example, for lakes of similar
528 surface area, even regionally proximate lakes may have vastly different amounts of benthic
529 habitat, mixing dynamics, and deep cold-water habitat. Deep, large lakes tend to have deeper
530 thermoclines and occasional metalimnetic dissolved oxygen peaks (Hanson et al. 2007).
531 Shallower lakes are also experiencing declines in mixing stability due to climate change
532 (Kraemer et al. 2015). Our model suggests there are many more shallow lakes than the area
533 scaling relationship suggests, implying more declines in mixing stability and shallower
534 thermoclines.
535
536 Our models indicate that concave lakes represent about 79% of all lakes in the conterminous
537 United States. Concave lakes are bowl-shaped, with deep water throughout the basin. The
538 opposite bathymetric shape, termed convex, describes lakes that are funnel-shaped with deep
539 water in a small portion of the basin (Figure 8 inset). However, as noted previously, under this
540 classification schema a perfectly conical lake ($V_d = 1$) is classified as moderately convex or

541 funnel shaped. The relative abundance of convex and concave lakes we estimated differs
542 substantially from previous work. For example, using bathymetric surveys Stachelek et al.
543 (2022) explored pattens in lake bathymetry for a sample of lakes available in LAGOS-US. We
544 calculate that the median volume development factor for lakes used in Stachelek et al. (2022)
545 was 1.2 whereas ours was 1.9. We attribute the higher number of concave lakes reported in
546 Stachelek et al. (2022) partially to the much greater prevalence of small lakes in our dataset
547 (417,442 < 10^5 m^2, 87%) compared to the bathymetric data in the cited study, which had
548 relatively few small lakes (562 < 10^5 m^2, 11%). However, this alone does not explain the entire
549 difference between our study and the work of Stachelek et al. (2022). For example, using a
550 subset of our modeled lakes that had a similar area distribution as those studied by Stachelek et
551 al. (2022), we estimated a volume development factor of 1.6. This is closer to the median value
552 reported by Stachelek et al. (2022) but still indicates that most lakes are concave or bowl shaped.
553 We also acknowledge, however, that in ~5% of lakes our models predicted mean depth greater
554 than the predicted maximum depth. This is an effect of fitting our models independently of one
555 another and indicated by $V_d > 3$. This small percentage of lakes occurred predominantly in the
556 smallest lakes (Figure 8). Our random forest model may overestimate mean depth in small lakes
557 (Supplemental Information Figure 4), inflating the volume development ratio, while in large
558 lakes this is not the case and reasonable ratios are predicted. However, we do note that our mean
559 depth model was more accurate than our maximum depth model (Table 3). Nevertheless, the

differences between our study and that of Stachelek et al. (2022) also highlight the importance of studying the full distribution of all lakes.

It should be noted that because we calculated the volume development ratio using both maximum and mean depth models (Equation 7), it was sensitive to uncertainty or bias in either model. As described above, our maximum depth model tended to overestimate the maximum depth in shallow lakes and underestimate the maximum depth in deep lakes, consistent with a low Hurst coefficient. Our mean depth model had a very similar Hurst coefficient, and biases were very similar (see also Figure 2 and Supplemental Information Figure 4). Thus, the bias in the volume development ratio, which would reflect differences in the bias between the maximum and mean depth models, is likely very small, providing some confidence in our estimates of the distribution of lake basin shapes across the conterminous United States.

For the purposes of modeling lake volumes, concave or bowl-shaped lakes will have larger volumes than predicted by a simple cone model, while convex or funnel-shaped lakes will have smaller volumes. The volumes we estimated for all lakes in the contiguous US (excluding the Laurentian Great Lakes) differed when calculated from our maximum depth model (1057 km^3) versus our mean depth model (1294 km^3). Our mean depth model consistently predicted higher lake volumes, and this is related to the fact that volume predictions made from the maximum depth model assumed a simple cone bathymetry, whereas the mean depth model does not. Despite differences in the total volume estimated between our maximum depth and mean depth models, our overall estimates for the total volume of lake water in the conterminous US were remarkably close to one another relative to the range calculated across other lake models

583 (Supplemental Information Table 1). Both our values were also similar to the overall mean
584 across all models (991 km^3).
585
586 In summary, our random forest models comprise the most extensive predictive modeling study of
587 lake depth to date. We have shown how satellite imagery may be used in conjunction with
588 traditional predictors of lake depth to improve depth estimates. However, our gains in predictive
589 power come at the cost of increased model complexity and care must be taken when applying our
590 depth estimates. The insights generated by our models provide new capability to understand a
591 range of lake attributes, including the frequency and distribution of lake depths and volumes,
592 which have important implications for the physics, chemistry, and biology of lakes.
593
594 Computer code used to generate the predictor variables and figures in this manuscript is available
595 at https://github.com/s-kganz/lakes/. This repository also includes serialized R objects of the
596 fitted models that can be loaded into an R session. Prediction results and detailed variable
597 importance metrics are available through the Environmental Data Initiative repository:
598 https://doi.org/10.6073/pasta/d68d5a218d1ec0351f8c3721d41e5e9c.

599 **5 Acknowledgements**

600 We thank the Rensselaer Institute for Data Exploration and Applications for access to computing
601 resources, the RPI Undergraduate Research Program, and US National Science Foundation grant
602 numbers 1754265 and 2048031. This work was also supported by the National Science
603 Foundation Graduate Research Fellowship Program under Grant No. DGE-2140004. Any
604 opinions, findings, and conclusions or recommendations expressed in this material are those of
605 the author(s) and do not necessarily reflect the views of the National Science Foundation. We

also acknowledge S. Oliver for guidance on previous modeling and K. Spence Cheruvelil and the LAGOS project for access to their lake depth data set. The authors declare no conflicts of interest.

6 Author Contributions

KJG: Data acquisition, analysis, drafting and editing the manuscript

MRG: Data acquisition, analysis, and editing of manuscript

KCR: Study design, drafting and editing the manuscript

615 **Tables**

616 Table 1. Categories of predictors included in our random forest models of lake depth. Detailed

617 information on predictors is given in Methods and Supplemental Information.

Category	Data sources	Number of initial candidate predictors	Number of predictors in Z_{max} model	Number of predictors in Z_{mean} model
Local terrain metrics	30m FABDEM; Messager et al. (2016)	49	4	4
Neighborhood terrain metrics	30m FABDEM	32	1	1
Water surface temperature	Landsat 8 TOA radiance; ERA5 monthly air temperature	12	6	4
Water surface reflectance	Landsat 8 surface reflectance	10	2	0
Lake polygon attributes	LAGOS-US	5	1	1
HUC4 summary statistics	LAGOS-US; 30m FABDEM	40	3	2
Other	LAGOS-US	17	1	1

618

Table 2. Summary of existing empirical models of maximum lake depth we evaluated in comparison to our model. Model performance metrics shown here are those reported in the cited studies. RMSE: root mean square error. R^2: proportion of variance explained. If multiple models are included in a particular study, the model with the best performance metrics (i.e., highest R^2, lowest RMSE) is described.

Citation	Region	Model	Predictors	R^2	RMSE
Oliver et al. (2016)	Northeastern USA	Linear mixed effects	Lake area, shoreline development index, nearshore slope, watershed to lake area ratio	0.29	7.1 m
Hollister et al. (2011)	Eastern USA	Linear regression	Nearshore slope, distance to lake center	0.67	5.09 m
Heathcote et al. (2015)	Southern Quebec	Linear regression	Nearshore elevation change	0.52	0.245 $\log_{10}$m
Sobek et al. (2011)	Sweden	Linear regression	Lake area, lake perimeter, nearshore slope	0.36	Not reported
Stachelek et al. (2022)	Northeastern and midwestern USA	Random forest	Bathymetric slope, distance to deepest point in lake	0.73	4.8 m
Becker and Daw (2005)	Wisconsin, USA	Linear regression	Nighttime water surface temperature	0.36	Not reported
Khazaei et al. (2022)	Global	Random forest	Lake area, lake perimeter, lake volume, watershed area, lake elevation	0.88	Not reported

Table 3. Model performance metrics for our maximum depth model (Z_{max}, top) and mean depth model (Z_{mean}, bottom) versus models from previous studies that we retrained using our lake depth dataset to enable objective comparison. We trained all models on depth observations in the LAGOS-US DEPTH module (Stachelek et al. 2021). RMSE: root mean square error; MAE: mean absolute error; PBIAS: percent bias; R^2: proportion of variance explained; H: median bootstrapped Hurst exponent from modeled Z_{max} depths of a similar area distribution to those in LAGOS-US DEPTH.

Model	Predictors (#)	RMSE (m)	MAE (m)	PBIAS (%)	R^2	H
Maximum depth						
Pruned random forest (this study)	18	8.04	5.02	-66	0.35	0.126
Oliver et al., (2016)	5	8.29	4.63	-29	0.17	0.236
Khazaei et al., (2022)	5	9.44	5.64	-63	0.17	0.141
Sobek (2011)	2	8.78	5.45	-65	0.15	1.046
Hollister et al., (2011)	2	8.18	5.11	-75	0.25	0.137
Heathcote et al., (2015)	1	9.72	5.49	-37	0.005	0.016
Mean depth						
Pruned random forest (this study)	13	2.97	1.89	-44	0.36	0.130
Khazaei et al., (2022)	5	3.22	2.07	-39	0.24	0.141
Messager et al. (2016)	2	3.14	2.06	-45	0.15	0.355

Figures

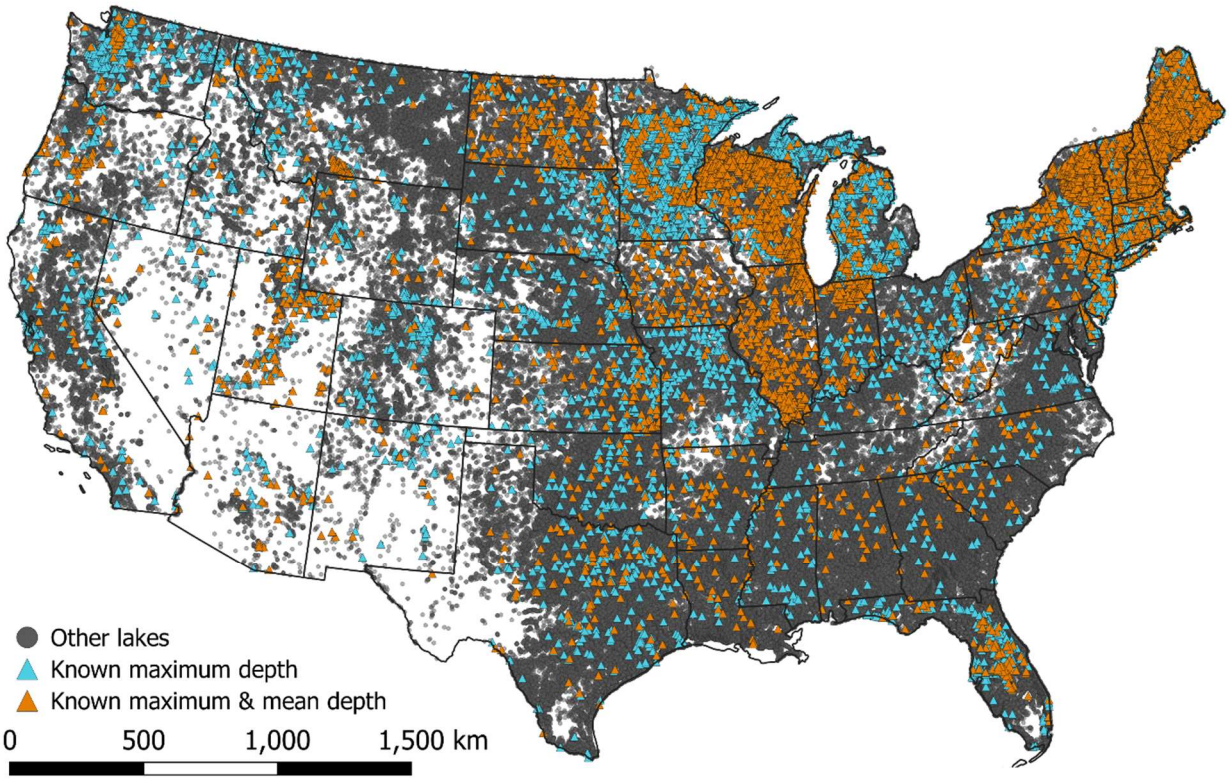


Figure 1: Map of lakes in the conterminous US, including those with known mean and maximum depth used to fit both our models (orange triangles, n = 5,881), those used to fit our maximum depth model only (blue triangles; n = 16,831), and all other lakes (gray circles, n = 476,833), Lakes larger than 1,000 ha with known depth(s) were excluded from modeling and do not appear in this map.

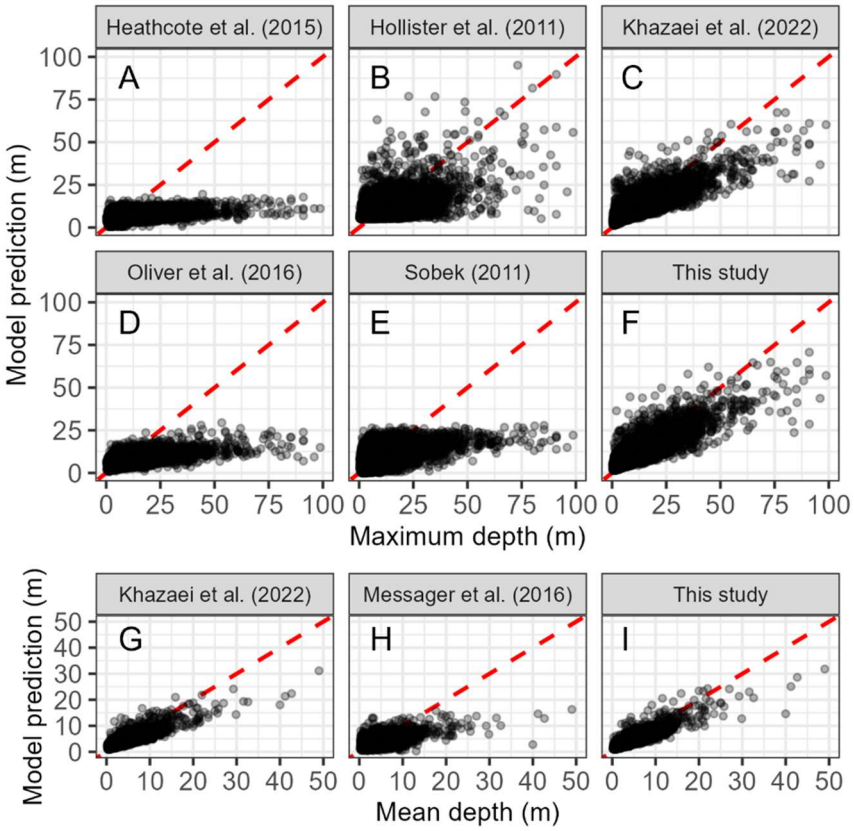


Figure 2: Model-predicted depths versus measured depths for the six maximum depth models (A-F) and three mean depth models (G-I) described in Table 3. Dashed red lines indicate 1:1 relationship between modeled and observed depths. Model performance metrics were derived from the nested resampling procedure described in our Methods, not from these fits. "This study" refers to our final pruned random forest models.

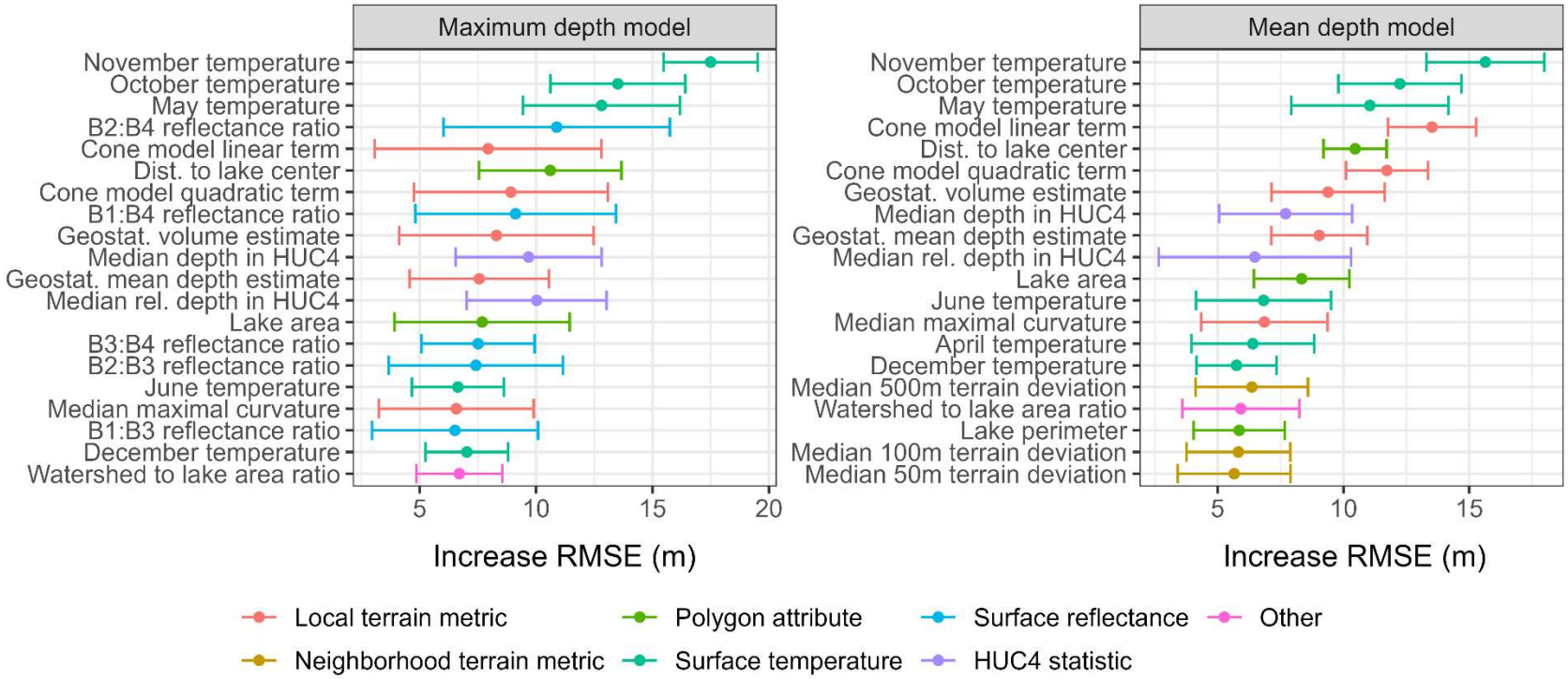


Figure 3: Aggregate permutation importance for the top 20 variables in the maximum depth (left) and mean depth (right) random forest models over 100 random forest fits via the Boruta algorithm. Error bars represent 1.68 sample standard deviations for the distribution of importance samples across all training repetitions. Full data for all variables in the random forest model is included in Supplemental Information.

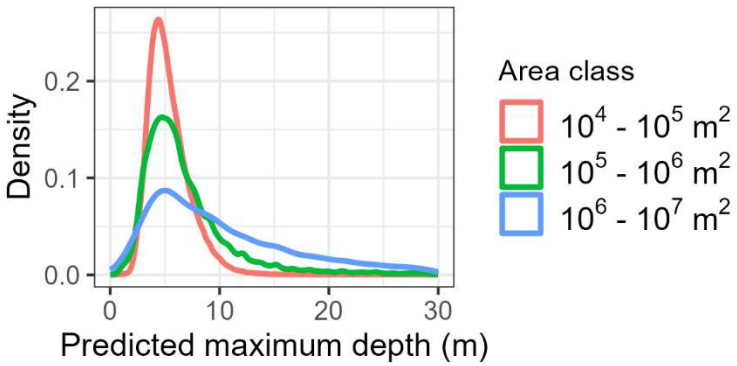
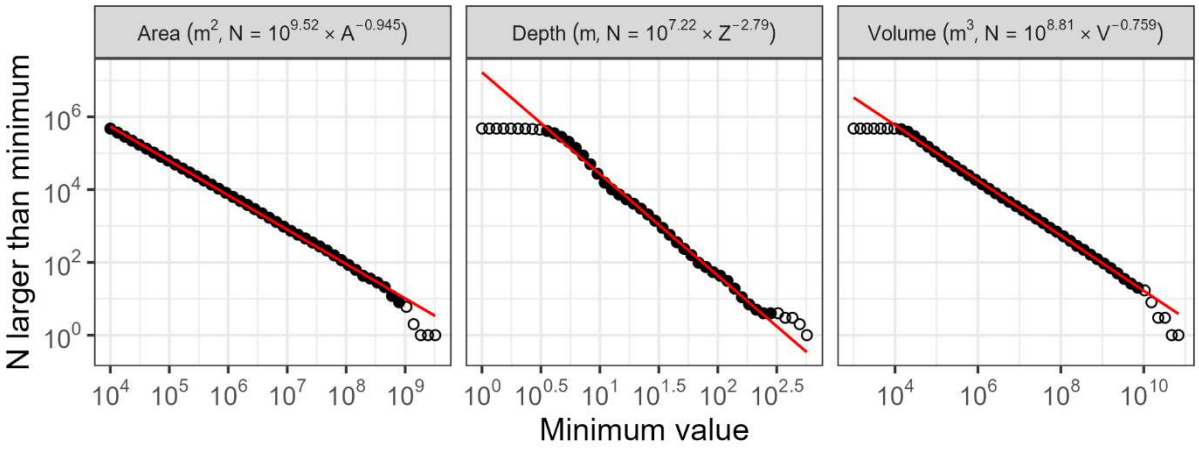
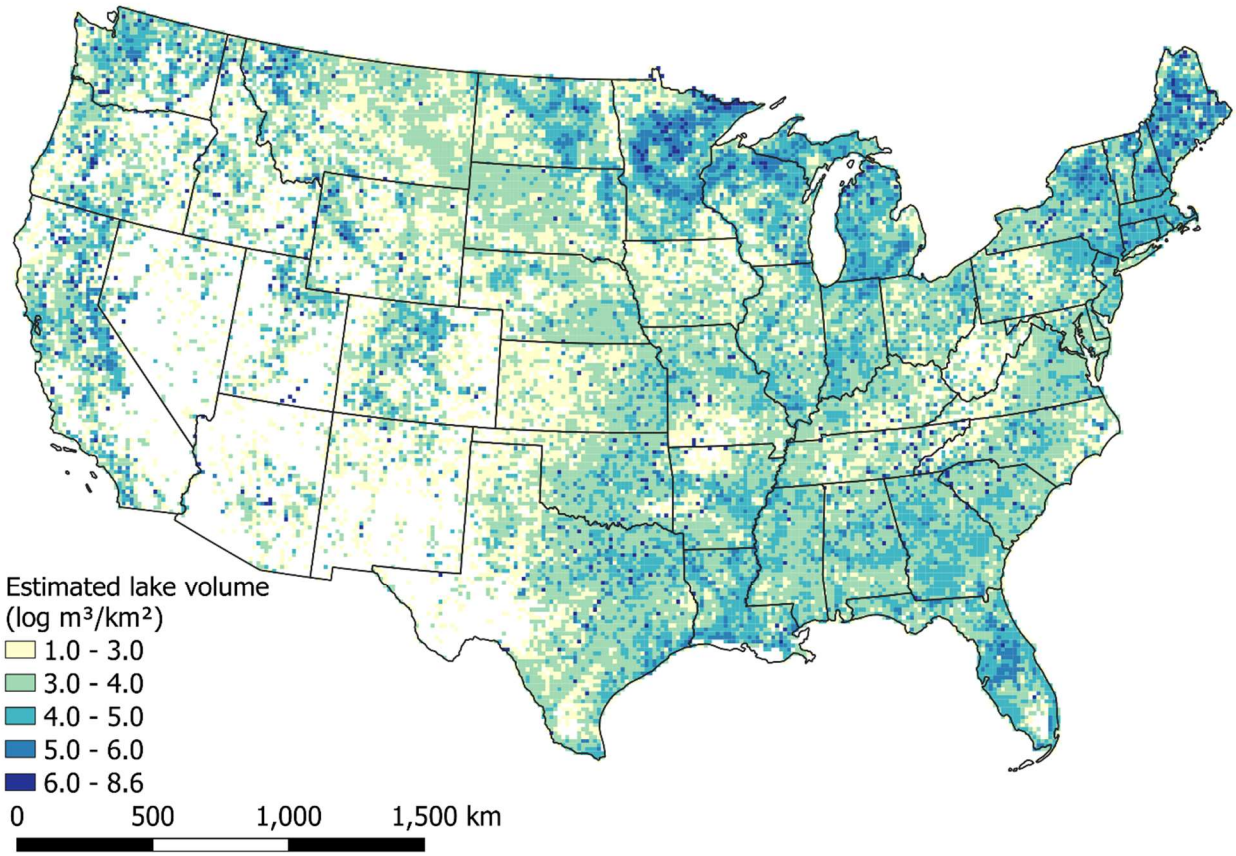


Figure 4: Depth density estimates for all modeled lakes in LAGOS-US smaller than 1,000 ha, stratified by lake area magnitude class. Mode depths for each area class are 4.37 m (10^4 – 10^5 m^2), 4.73 m (10^5 – 10^6 m^2), and 4.96 m (10^6 – 10^7 m^2). Note that lakes deeper than 30 m were present, but the plot is truncated to better visualize differences in the distributions.



659

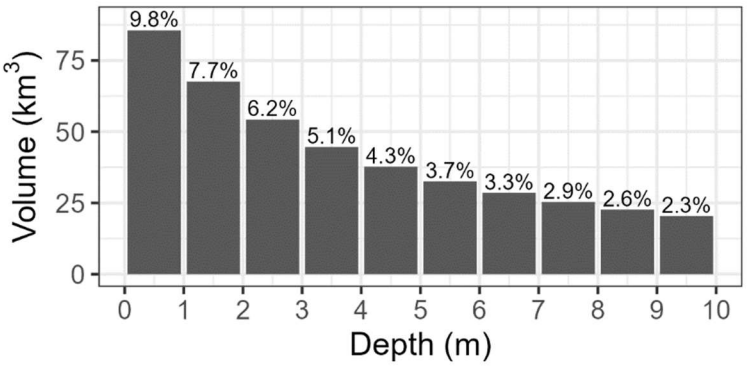
660 Figure 5: Regression (red line) and data (points) for lake maximum depth, area, and volume in
661 LAGOS-US, modeled after a Pareto distribution model described in Messager et al. (2016). The
662 regression was fit using both modeled and observed lake depths for lakes with area < 1,000 ha
663 and observed depths for lakes with area ≥ 1,000 ha. If a lake was used for model training or
664 validation, the observed depth was used over the modeled depth. Filled circles are included in the
665 regression, while open circles are excluded to reduce the influence of few large lakes and the
666 minimum area limit (1 ha) in the LAGOS-US dataset. The data ranges used for each regression
667 are $(10^4, 10^9)$ m^2, $(10^{0.5}, 10^{2.5})$, and $(10^4, 10^{10})$ m^3.



668

669 Figure 6: Water volume per unit area estimate derived from the cone model described in

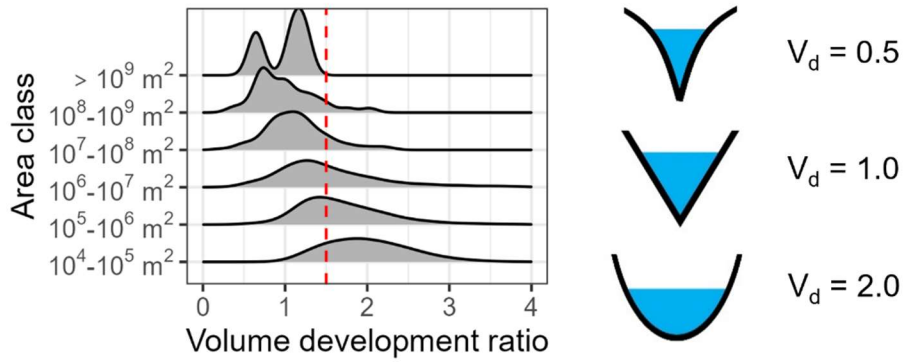
670 Equation 2 and Z_{max} chosen as in Figure 5.



671

672 Figure 7: Water volumes contained within 1-meter intervals of selected lakes chosen as in Figure
673 5, assuming a conical basin geometry. The label above each bar indicates the proportion of all lake
674 water in that depth interval. Each bar is plotted at the center of a 1-meter interval. For example,
675 the first bar corresponds to the interval [0 m, 1 m], the second to [1 m, 2 m], etc.

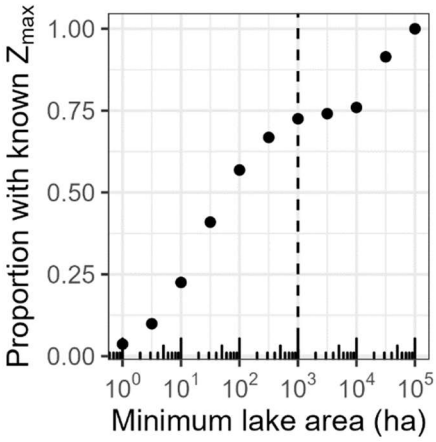
676



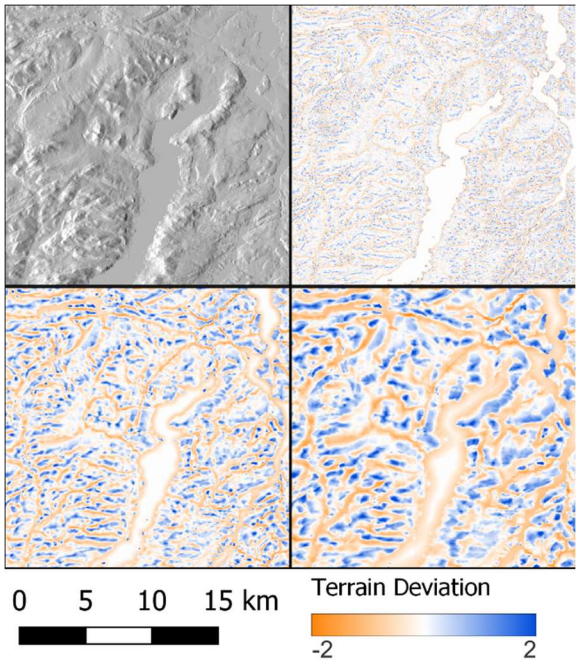
677

678 Figure 8: Normalized relative density of volume development ratio for modeled and observed
679 lakes, faceted by lake area. Vertical dashed line at $V_d = 1.5$ is the border between convex and
680 concave lakes as defined by hypsometry. Note that the bimodal distribution in the largest size
681 class is an artifact from that category having only 6 lakes. Diagrams at right indicate approximate
682 basin shape for different volume development ratios.

7 Supplemental Information

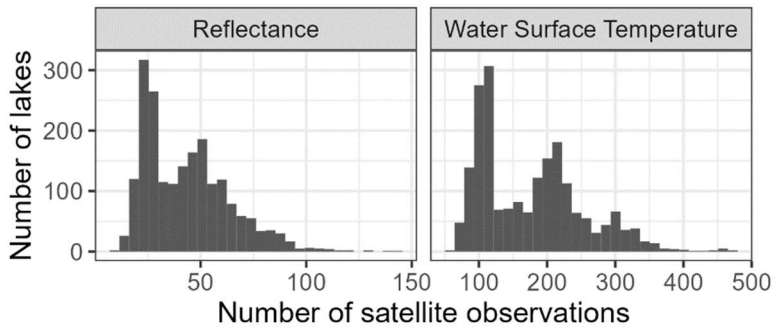


Supplemental Information Figure 1: Proportion of all lakes in the conterminous US with known maximum depth as reported in the LAGOS-US DEPTH module, sorted by area. Vertical dotted line represents the maximum lake area included for modeling in this study.

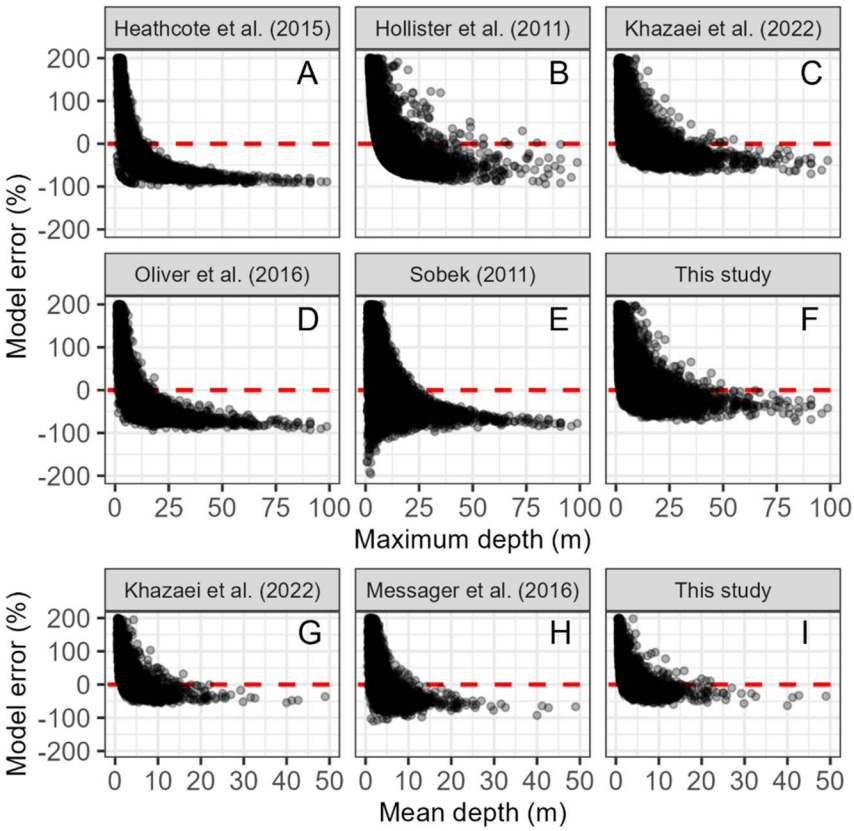


690

691 Supplemental Information Figure 2: Hillshade and terrain deviation rasters of varying kernel
692 sizes near Lake George, New York, USA. Negative values (orange) indicate valleys, while
693 positive values (blue) indicate ridges. As kernel size increases, deviation metrics correspond to
694 larger topographic features. Upper left: hillshade of 30 m FABDEM with illumination at 270°
695 azimuth and 45° elevation. Upper right: 100m radius kernel. Lower left: 500 m radius kernel.
696 Lower right: 1,000 m radius kernel.



Supplemental Information Figure 3: Number of satellite observations associated with surface reflectance (left) and water surface temperature (right) available in a random sample of 2,000 lakes from the LAGOS-US LOCUS module. Peaks in the distribution result from Landsat 8 orbit overlaps.

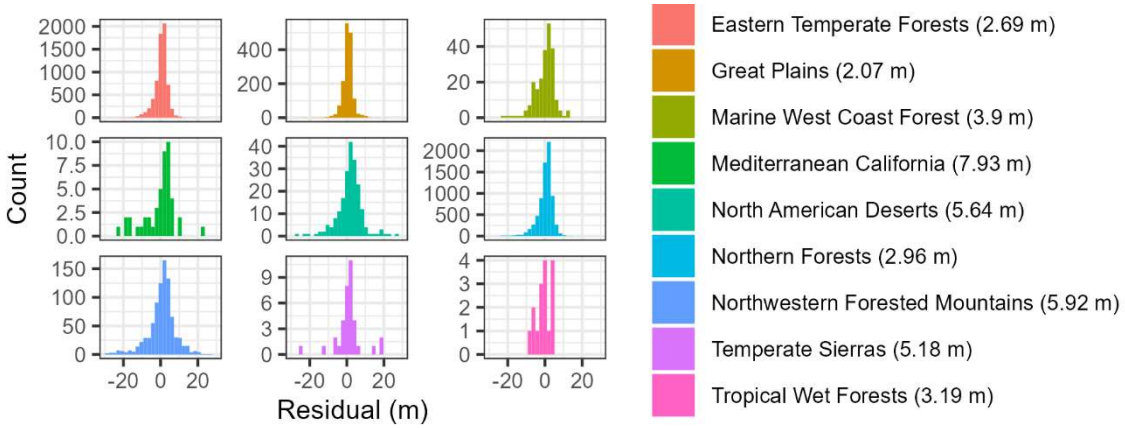


703

704 Supplemental Information Figure 4: As in Figure 2, but with prediction error instead of

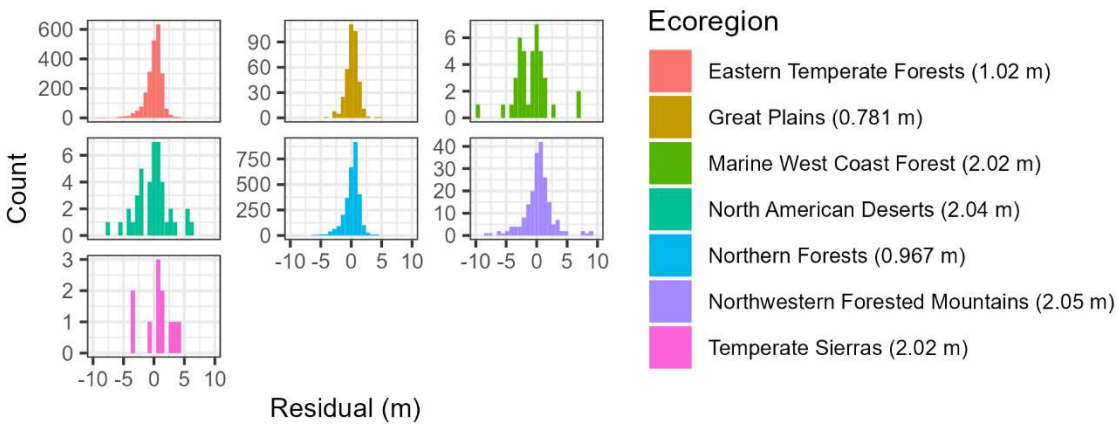
705 prediction value. The dotted red line is equivalent to that in Figure 2 i.e. indicating perfect

706 agreement between modeled and actual depths.



707

708 Supplemental Information Figure 5: Residuals for the final random forest maximum depth
709 model, faceted by US Environmental Protection Agency Level 1 Ecoregions (Omernik and
710 Griffith 2014). Only ecoregions with at least 10 observations are included (note varying y-axis
711 limits). Median absolute residual for each ecoregion is reported in parentheses in the figure
712 legend.



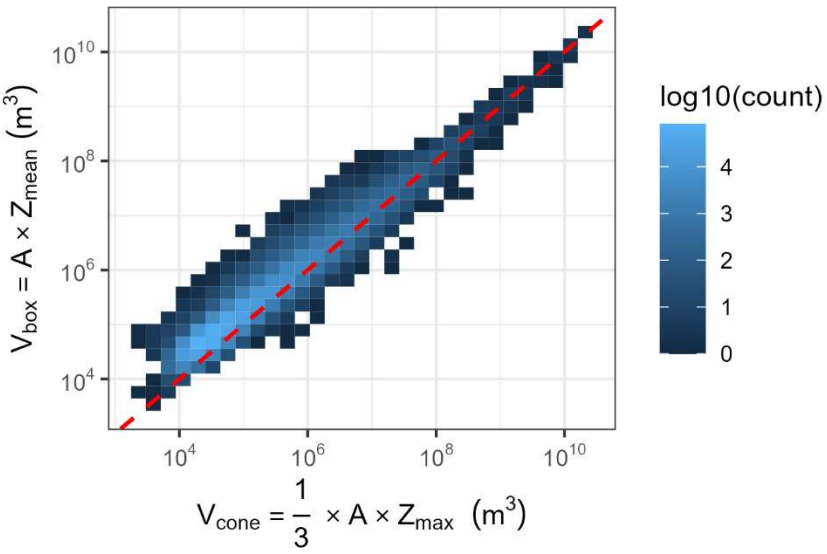
713

714 Supplemental Information Figure 6: As in Supplemental Information Figure 5, except using

715 mean depth predictions instead of maximum depth predictions. Fewer ecoregions are present

716 because of the fewer number of mean depth observations overall.

717



718

719 Supplemental Information Figure 7. Volume estimates calculated from our mean depth model
720 were closely correlated with volume estimates calculated from our maximum depth model ($R^2 =$
721 0.55), although volumes estimated using the mean depth model were consistently higher. Red
722 line is 1:1.

723

Predicting lake depth and volume using all lakes, including those > 1,000 ha

In the manuscript we modeled lake depth for lakes < 1,000 ha in area because extremely large lakes are disproportionately likely to have known depth (see Supplemental Information Figure 1) and are typically outliers with respect to depth (Sobek et al. 2011; Oliver et al. 2016). However, predicting the total volume of all US lakes necessitates estimating lake depth and volume for the roughly 20% of lakes > 1,000 ha whose depths are not known. To do this, we reran our maximum and mean depth models following the procedures outlined in our methods section including lakes > 1,000 ha (but again excluding the Laurentian Great Lakes). The results of this procedure are shown in Supplemental Information Table 1 below.

Supplemental Information Table 1. Model performance metrics for our maximum depth model (Z_{max}, top) and mean depth model (Z_{mean}, bottom) for all lakes including those > 1,000 ha (but excluding the Laurentian Great Lakes) versus selected other comparable models from published studies. We trained all models on depth observations in the LAGOS-US DEPTH module (Stachelek et al. 2021). RMSE: root mean square error; MAE: mean absolute error; PBIAS: percent bias; R^2: proportion of variance explained; H: median bootstrapped Hurst exponent from modeled Z_{max} depths of a similar area distribution to those in LAGOS-US DEPTH; Volume: the total volume of all lakes in the conterminous US estimated using this model.

Model	Predictors (#)	RMSE (m)	MAE (m)	PBIAS (%)	R^2	H	Volume (km^3)
Maximum depth							
Pruned random forest	13	12.0	5.90	-67	0.38	0.143	1,057
Oliver et al., (2016)	5	11.6	5.13	-29	0.23	0.275	513
Khazaei et al., (2022)	5	13.8	6.76	-69	0.18	0.241	1,138
Sobek (2011)	2	12.3	6.74	-62	0.17	0.736	747
Hollister et al., (2011)	2	10.2	5.87	-80	0.32	0.119	1,010
Heathcote et al., (2015)	1	14.1	6.75	-33	0.007	0.018	286
Mean depth							
Pruned random forest	9	2.89	1.73	-42	0.53	0.150	1,294
Khazaei et al., (2022)	5	4.24	2.39	-42	0.33	0.157	1,616
Messager et al. (2016)	2	4.27	2.62	-61	0.06	0.020	1,254

Conic section volume derivation

For a cone with height H and radius R, the volume of a section of the cone may be defined as:

$$V_{section} = \int_{h_1}^{h_2} \pi r(h)^2 dh \, , h_1 < h_2 < H$$

Where $r(h)$ is the radius of a horizontal slice of the cone at height h above the base. The ratio of the radius to the distance to the point of the cone is constant for all horizontal slices of the cone:

$$\frac{r(h)}{H-h} = \frac{R}{H}, h < H$$

So:

$$r(h) = \frac{R(H-h)}{H}$$

Substituting into the integral yields:

$$V_{section} = \int_{h_1}^{h_2} \pi \frac{R^2(H-h)^2}{H^2} dh$$

$$= \frac{\pi R^2}{3H^2}[(H-h_1)^3 - (H-h_2)^3]$$

The equation may be put in terms of known variables by recognizing πR^2 is equivalent to lake area and H is equivalent to Z_{max}. Changing h_i to z_i, we arrive at the following equation:

$$V_{section} = \frac{A}{3Z_{max}{}^2}[(Z_{max}-z_1)^3 - (Z_{max}-z_2)^3], z_1 < z_2 < Z_{max}$$

Cael et al. (2017) lake volume comparison

Supplemental Information Table 2. We calculated a comparable estimate of the total volume of lakes modeled in this study from Table 2 in Cael et al. (2017). This table lists average volume for lakes by size class. Assuming that all lakes in our dataset had the appropriate average volume given in the table, we computed total volume and mean depth using area data in LAGOS-US LOCUS as in the table below.

Lake area range (m^2)	N with modeled or observed depth data in LAGOS US	Average lake volume (m^3)	Est. total volume (m^3)	Total area (m^2)	Mean depth (m)
$10^4 – 10^5$	417,442	7.50×10^4	3.13×10^{10}	1.11×10^{10}	2.83
$10^5 – 10^6$	52,725	1.06×10^6	5.59×10^{10}	1.42×10^{10}	3.95
$10^6 – 10^7$	6,676	1.53×10^7	1.02×10^{11}	1.78×10^{10}	5.68
$10^7 – 10^8$	991	2.61×10^8	2.59×10^{11}	2.00×10^{10}	9.42
$10^8 – 10^9$	120	4.16×10^9	4.99×10^{11}	2.33×10^{10}	16.5
$>10^9$	6	6.73×10^{10}	4.04×10^{11}	9.85×10^9	41.0
Totals	**477,960**	**NA**	**1.35×10^{12}**	**9.63×10^{10}**	**14.0**

8 References

Becker, M. W., and A. Daw. 2005. Influence of lake morphology and clarity on water surface temperature as measured by EOS ASTER. Remote Sens. Environ. **99**: 288–294. doi:10.1016/j.rse.2005.09.003

Birge, E. A. 1915. The heat budgets of American and European lakes,.

Breiman, L. 2001. Random Forests. Mach. Learn. **45**: 5–32. doi:10.1023/A:1010933404324

Brenning, A. 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. *2012 IEEE International Geoscience and Remote Sensing Symposium*. Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium. 5372–5375.

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

Cael, B. B., and D. Seekell. 2022. The relationship between lake surface area and maximum depth.

Calamita, E., S. Piccolroaz, B. Majone, and M. Toffolon. 2021. On the role of local depth and latitude on surface warming heterogeneity in the Laurentian Great Lakes. Inland Waters **11**: 208–222. doi:10.1080/20442041.2021.1873698

Carpenter, S. R. 1983. Lake geometry: Implications for production and sediment accretion rates. J. Theor. Biol. **105**: 273–286. doi:10.1016/S0022-5193(83)80008-3

Cheruvelil, K. S., P. A. Soranno, I. M. McCullough, K. E. Webster, L. K. Rodriguez, and N. J. Smith. 2021. LAGOS-US LOCUS v1.0: Data module of location, identifiers, and

physical characteristics of lakes and their watersheds in the conterminous U.S. Limnol. Oceanogr. Lett. **6**: 270–292. doi:10.1002/lol2.10203

Dodds, P. S., and D. H. Rothman. 2000. Scaling, Universality, and Geomorphology. Annu. Rev. Earth Planet. Sci. **28**: 571–610. doi:10.1146/annurev.earth.28.1.571

Douglas, D. H., and T. K. Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartogr. Int. J. Geogr. Inf. Geovisualization **10**: 112–122. doi:10.3138/FM57-6770-U75U-7727

Downing, J. A., Y. T. Prairie, J. J. Cole, 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

Du, C., H. Ren, Q. Qin, J. Meng, and S. Zhao. 2015. A Practical Split-Window Algorithm for Estimating Land Surface Temperature from Landsat 8 Data. Remote Sens. **7**: 647–665. doi:10.3390/rs70100647

Efremova, T. V., and N. I. Pal'shin. 2011. Ice phenomena terms on the water bodies of Northwestern Russia. Russ. Meteorol. Hydrol. **36**: 559–565. doi:10.3103/S1068373911080085

Garcia-Castellanos, D., and U. Lombardo. 2007. Poles of inaccessibility: A calculation algorithm for the remotest places on earth. Scott. Geogr. J. **123**: 227–233. doi:10.1080/14702540801897809

Geyman, E. C., and A. C. Maloof. 2019. A Simple Method for Extracting Water Depth From Multispectral Satellite Imagery in Regions of Variable Bottom Type. Earth Space Sci. **6**: 527–537. doi:10.1029/2018EA000539

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 10.

Håkanson, L. 1977. On Lake Form, Lake Volume and Lake Hypsographic Survey. Geogr. Ann. Ser. Phys. Geogr. **59**: 1–30. doi:10.2307/520579

Håkanson, L. 1981. A manual of lake morphometry, Springer Berlin Heidelberg.

Håkanson, L. 2004. Lakes: form and function, 1st ed. Blackburn Press.

Håkanson, L., and B. Karlsson. 1984. On the Relationship between Regional Geomorphology and Lake Morphometry. A Swedish Example. Geogr. Ann. Ser. Phys. Geogr. **66**: 103–119. doi:10.2307/520942

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

Hawker, L., P. Uhe, L. Paulo, J. Sosa, J. Savage, C. Sampson, and J. Neal. 2022. A 30 m global map of elevation with forests and buildings removed. Environ. Res. Lett. **17**: 024016. doi:10.1088/1748-9326/ac4d4f

Heathcote, A. J., P. A. del Giorgio, and Y. T. Prairie. 2015. Predicting bathymetric features of lakes from the topography of their surrounding landscape D. Brickman [ed.]. Can. J. Fish. Aquat. Sci. **72**: 643–650. doi:10.1139/cjfas-2014-0392

Hutchinson, G. 1957. A Treatise on Limnology, Geography, Physics and Chemistry, John Wiley and Sons.

Jane, S. F., G. J. A. Hansen, B. M. Kraemer, and others. 2021. Widespread deoxygenation of temperate lakes. Nature **594**: 66–70. doi:10.1038/s41586-021-03550-y

836 Khazaei, B., L. K. Read, M. Casali, K. M. Sampson, and D. N. Yates. 2022. GLOBathy, the
837 global lakes bathymetry dataset. Sci. Data **9**: 36. doi:10.1038/s41597-022-01132-9
838 Kirillin, G., M. Leppäranta, A. Terzhevik, and others. 2012. Physics of seasonally ice-covered
839 lakes: a review. Aquat. Sci. **74**: 659–682. doi:10.1007/s00027-012-0279-y
840 Kraemer, B. M., O. Anneville, S. Chandra, and others. 2015. Morphometry and average
841 temperature affect lake stratification responses to climate change. Geophys. Res. Lett. **42**:
842 4981–4988. doi:10.1002/2015GL064097
843 Kursa, M. B., and W. R. Rudnicki. 2010. Feature Selection with the Boruta Package. J. Stat.
844 Softw. **36**: 1–13. doi:10.18637/jss.v036.i11
845 Lang, M., M. Binder, J. Richter, and others. 2019. mlr3: A modern object-oriented machine
846 learning framework in R. J. Open Source Softw. doi:10.21105/joss.01903
847 Lathrop, R. G. 1992. Landsat Thematic Mapper monitoring of turbid inland water quality.
848 Photogramm. Eng. Remote Sens. **58**: 465–470.
849 Lovelace, R., J. Nowosad, and J. Muenchow. 2022. Statistical Learning, *In* Geocomputation with
850 R.
851 Matheson, D. H., and M. Munawar. 1978. Lake Superior Basin and its Development. J. Gt.
852 Lakes Res. **4**: 249–263. doi:10.1016/S0380-1330(78)72196-9
853 Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume
854 and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. **7**:
855 13603. doi:10.1038/ncomms13603
856 Meyer, H., and E. Pebesma. 2022. Machine learning-based global maps of ecological variables
857 and the challenge of assessing them. Nat. Commun. **13**: 2208. doi:10.1038/s41467-022-
858 29838-9

Muñoz Sabater, J. 2018. ERA5 hourly data on single levels from 1979 to present.

Oliver, S. K., P. A. Soranno, C. E. Fergus, and others. 2016. Prediction of lake depth across a 17-state region in the United States. Inland Waters **6**: 314–324. doi:10.1080/IW-6.3.957

Olmanson, L. G., P. L. Brezonik, J. C. Finlay, and M. E. Bauer. 2016. Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. Remote Sens. Environ. **185**: 119–128. doi:10.1016/j.rse.2016.01.007

Omernik, J. M., and G. E. Griffith. 2014. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. Environ. Manage. **54**: 1249–1266. doi:10.1007/s00267-014-0364-1

Pinsky, M. L., B. Worm, M. J. Fogarty, J. L. Sarmiento, and S. A. Levin. 2013. Marine Taxa Track Local Climate Velocities. Science **341**: 1239–1242. doi:10.1126/science.1239352

QGIS Development Team. 2021. QGIS Geograhpic Information System. Open Source Geospatial Foundation.

R Core Team. 2021. R: A language and environment for statistical computing.

Read, E. K., V. P. Patil, S. K. Oliver, and others. 2015. The importance of lake-specific characteristics for water quality across the continental United States. Ecol. Appl. Publ. Ecol. Soc. Am. **25**: 943–955. doi:10.1890/14-0935.1

Roy, D. P., M. A. Wulder, T. R. Loveland, and others. 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens. Environ. **145**: 154–172. doi:10.1016/j.rse.2014.02.001

Safanelli, J. L., R. R. Poppiel, L. F. C. Ruiz, B. R. Bonfatti, F. A. de O. Mello, R. Rizzo, and J. A. M. Demattê. 2020. Terrain Analysis in Google Earth Engine: A Method Adapted for

High-Performance Global-Scale Analysis. ISPRS Int. J. Geo-Inf. **9**: 400. doi:10.3390/ijgi9060400

Seekell, D. A., and M. L. Pace. 2011. Does the Pareto distribution adequately describe the size-distribution of lakes? Limnol. Oceanogr. **56**: 350–356. doi:10.4319/lo.2011.56.1.0350

Sobek, S., J. Nisell, and J. Fölster. 2011. Predicting the depth and volume of lakes from map-derived parameters. Inland Waters **1**: 177–184. doi:10.5268/IW-1.3.426

Stachelek, J., P. J. Hanly, and P. A. Soranno. 2022. Imperfect slope measurements drive overestimation in a geometric cone model of lake and reservoir depth. Inland Waters. doi:10.1080/20442041.2021.2006553

Stachelek, J., L. K. Rodriguez, J. Díaz Vázquez, and others. 2021. LAGOS-US DEPTH v1.0: Data module of observed maximum and mean lake depths for a subset of lakes in the conterminous U.S.doi:10.6073/PASTA/64DDC4D04661D9AEF4BD702DC5D8984F

Stanley, E. H., S. M. Collins, N. R. Lottig, S. K. Oliver, K. E. Webster, K. S. Cheruvelil, and P. A. Soranno. 2019. Biases in lake water quality sampling and implications for macroscale research. Limnol. Oceanogr. **64**: 1572–1585. doi:10.1002/lno.11136

Stumpf, R. P., K. Holderied, and M. Sinclair. 2003. Determination of water depth with high-resolution satellite imagery over variable bottom types. Limnol. Oceanogr. **48**: 547–556. doi:10.4319/lo.2003.48.1_part_2.0547

Taranu, Z. E., and I. Gregory-Eaves. 2008. Quantifying Relationships Among Phosphorus, Agriculture, and Lake Depth at an Inter-Regional Scale. Ecosystems **11**: 715–725. doi:10.1007/s10021-008-9153-0

Vanhellemont, Q. 2020. Automated water surface temperature retrieval from Landsat 8/TIRS. Remote Sens. Environ. **237**: 111518. doi:10.1016/j.rse.2019.111518

Walsh, S. E., S. J. Vavrus, J. A. Foley, V. A. Fisher, R. H. Wynne, and J. D. Lenters. 1998. Global patterns of lake ice phenology and climate: Model simulations and observations. J. Geophys. Res. Atmospheres **103**: 28825–28837. doi:10.1029/98JD02275

Webster, K. E., I. M. McCullough, and P. A. Soranno. 2022. Deeper by the Dozen: Diving into a Database of 17,675 Depths for U.S. Lakes and Reservoirs. Limnol. Oceanogr. Bull. **31**: 1–5. doi:10.1002/lob.10482

West, W. E., K. P. Creamer, and S. E. Jones. 2016. Productivity and depth regulate lake contributions to atmospheric methane. Limnol. Oceanogr. **61**: S51–S61. doi:10.1002/lno.10247

Wetzel, R. G., and G. E. Likens. 2000. Limnological analyses, Springer New York.

Wright, M. N., and A. Ziegler. 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. **77**: 1–17. doi:10.18637/jss.v077.i01

Zhan, P., C. Song, K. Liu, T. Chen, L. Ke, S. Luo, and C. Fan. 2023. Can we estimate the lake mean depth and volume from the deepest record and auxiliary geospatial parameters? Journal of Hydrology **617**: 128958. doi:10.1016/j.jhydrol.2022.128958