

Research Update: 25 October 2021

Keenan Ganz

This notebook summarizes the work I have done on the lake depth modeling project so far.

Literature Review

There are two main lake depth modeling strategies in the literature that differ in their predictors. In the first, characteristics of the terrain in the lake basin are extended below the water line to estimate maximum water depth. In the second, other characteristics of a lake - including shoreline geometry, watershed size, and position in the drainage network - are correlated with existing maximum depth measurements.

Both of these strategies are represented in R packages as well. `lakemorpho` implements the basin morphology approach while `LAGOSNE` implements the correlative approach alongside an interface to other lake data^{1,2}.

In its simplest form, the basin morphology modeling approach computes the median slope in a lake basin and assumes that lake bathymetry has a similar shape. Estimated lake depth, denoted \hat{Z} , is calculated as

$$\hat{Z} = D \times S_1 \tag{1}$$

where D is the maximum distance from shoreline and S_1 is the median slope within the lake basin³. Few published studies opt for one depth modeling technique or the other. Several topographic and other variables are combined to produce the best model for the region under study. Citations, model selection, predictors, and model performance for studies published within the last decade are shown in Table 1.

There is some variation in how the lake basin polygon is determined among published literature. Intuitively, larger lakes probably have larger basins. So, a larger buffer around a lake polygon is necessary to capture information on the surrounding topography. At the same time, too large a buffer may extend beyond the lake basin and add superfluous topography to the slope calculation. Several basin calculation methods are present in the literature, including a constant buffer distance, a buffer distance dependent on lake size, and buffers incorporating drainage information.

Table 1: Summary of existing models of maximum lake depth, using local topography and regional characteristics. \hat{Z} : maximum lake depth; D : maximum distance from shore in lake; S_{median} : median slope in lake catchment.

Citation	Region	Model	Most Important Predictor	Maximum Depth R^2	RMSE
Oliver et al. (2016)	Northeastern United States	Mixed effects (regional and observational)	Not reported	0.29	7.1m
Hollister et al. (2011)	NHD HUC 01 and 02	$\hat{Z} \approx D \cdot S_{median}$	Median slope	0.48-0.67	5.09-5.95m
Heathcote et al. (2015)	Southern Quebec, Canada	Log-transformed linear regression	Elevation change 25m from lake	0.52	0.245 \log_{10} m
Sobek (2011)	Sweden	Log-transformed linear regression and partial least squares regression	Lake area, perimeter, median slope	0.36	Not reported

Direct comparison of published models is impeded by inconsistent study region data sources. The variability of the underlying lake data clearly impacts the predictive power of the resulting model. The two studies with the highest reported R^2 also had the most restrictive study region. However, some patterns are evident. Although topographic information alone accounts for <50% of the variability in maximum depth, topographic variables are represented in all four models in Table 1. Other lake characteristics have some predictive power, but terrain metrics are an essential part of lake depth modeling.

The goals of this project have been to recreate the above models using a common dataset, explore the predictive power of other terrain metrics, consider the predictive power of the geologic context of a lake basin, and to compare the effect of buffer procedure on predicted lake depth.

Methods

Data Sources

Surveyed maximum lake depths and lake polygons were acquired from the EPA National Lakes Assessment, carried out in 2007 and 2012⁴. For lakes that were sampled in both campaigns, only the most recent survey data was retained. Topographic data was computed from the Shuttle Radar Topography Mission 30m digital elevation model.⁵ Geologic units were identified from the state geologic map compilation⁶.

Spatial Processing

Spatial operations were carried out in QGIS. Large rasters were processed on the cloud GIS platform Google Earth Engine (GEE)⁷. The general workflow was as follows:

1. Compute buffer regions around lake polygons.
2. Upload buffer regions to GEE.
3. Compute zonal statistics for a given raster dataset within each buffer region.
4. Perform statistical analysis in R⁸.

Specific processing steps are described in accompanying notebooks.

Results

Existing Model Comparison

The four models in Table 1 were recreated as closely as possible for the present data sources. Some parameters, such as length scale or buffer size, were adjusted to account for the precision of existing data.

Table 2: Model performance metrics for existing lake depth models when applied to National Lakes Assessment survey data.

Citation	RMSE (m)	R^2
Hollister et al. (2011)	7.50	0.27
Heathcote et al. (2015)	8.10	0.24
Sobek (2011)	7.04	0.36

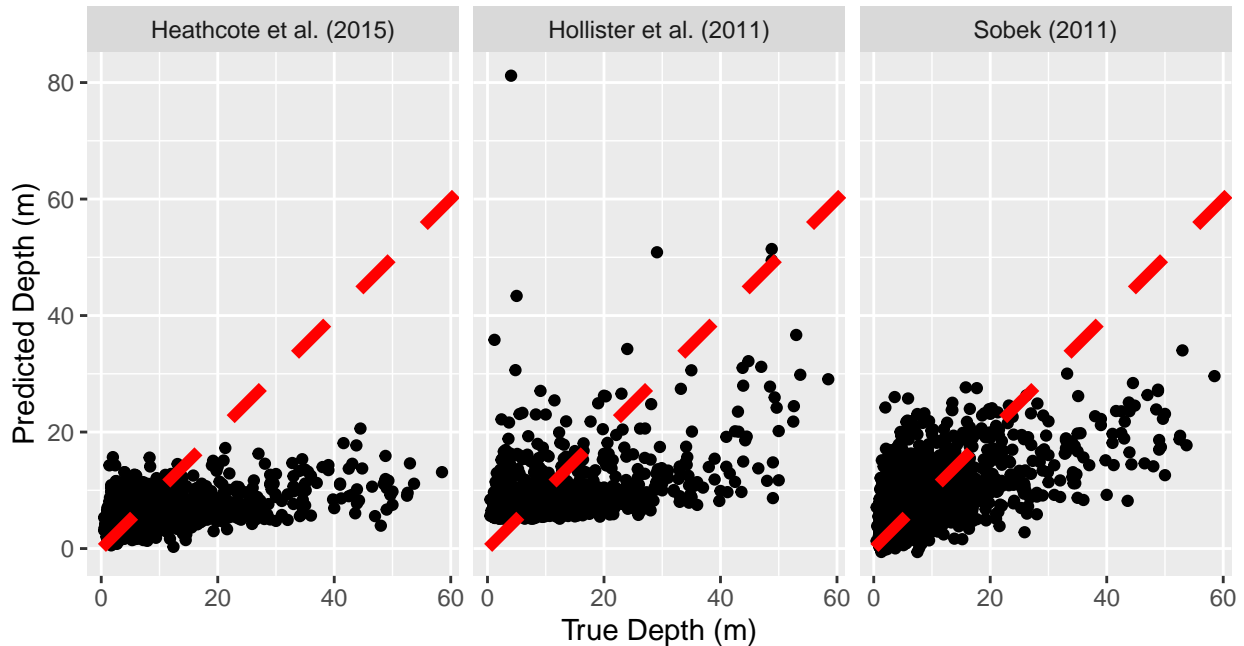


Figure 1: Scatterplots of predicted and actual maximum lake depths under the modeling regimes described in Table 1. Dotted lines indicate 1:1 correspondence.

In all three cases tested, the model performs worse on the NLA dataset than on the dataset from the source publication. This is probably due to the wider spatial extend and more varied environmental context of this data.

Other Terrain Metrics

Slope is not the only descriptor of terrain morphology. We also considered profile curvature, which is the second derivative of elevation in the direction of maximum slope. We add a quadratic term to the basin morphology model,

$$\hat{Z} = D \times S_1 + \frac{D^2}{2} \times S_2 \quad (2)$$

where S_2 is profile curvature.

The most immediate comparison for this model is Hollister et al. The linear and quadratic terms are somewhat correlated ($r = 0.66$). The quadratic term alone accounts for roughly 20% of the variance. But, the linear and quadratic terms together only account for 35% of the variance.

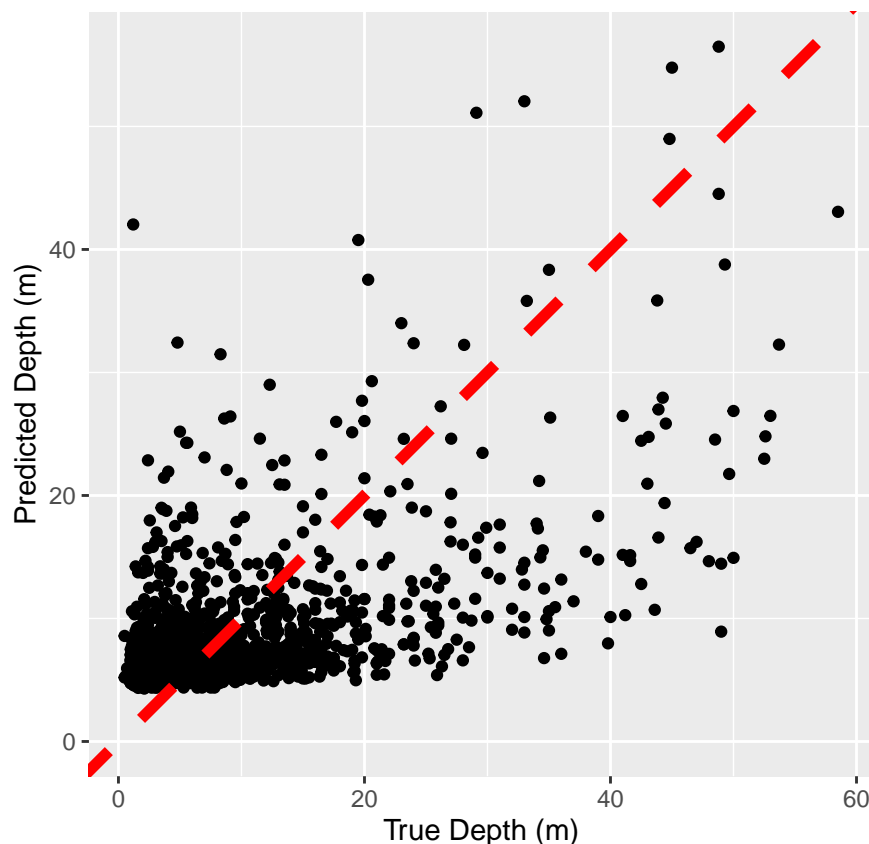


Figure 2: Predicted an actual maximum lake depth using Eq. 2. Dashed line indicates 1:1 correspondence.

Other terrain metrics are also candidate predictors. Metrics calculated from a moving window, such as relief and deviation, are complicated by adding another parameter to set the size of the moving window. The choice of parameter determines the scale of topographic feature captured.

Geologic Context of Lake Basins

The geologic context of lake basins may be related to sedimentation rate and bathymetry. This, in turn, may make geologic context a predictor of lake depth. We tested a dataset of surficial sediments and underlying bedrock. In both cases, predictive power was low ($R^2 < 0.1$). However, co-occurrence analysis revealed a

pattern of note in surficial sediments. A probabilistic model⁹ was applied to identify sediment types that positively co-occurred.

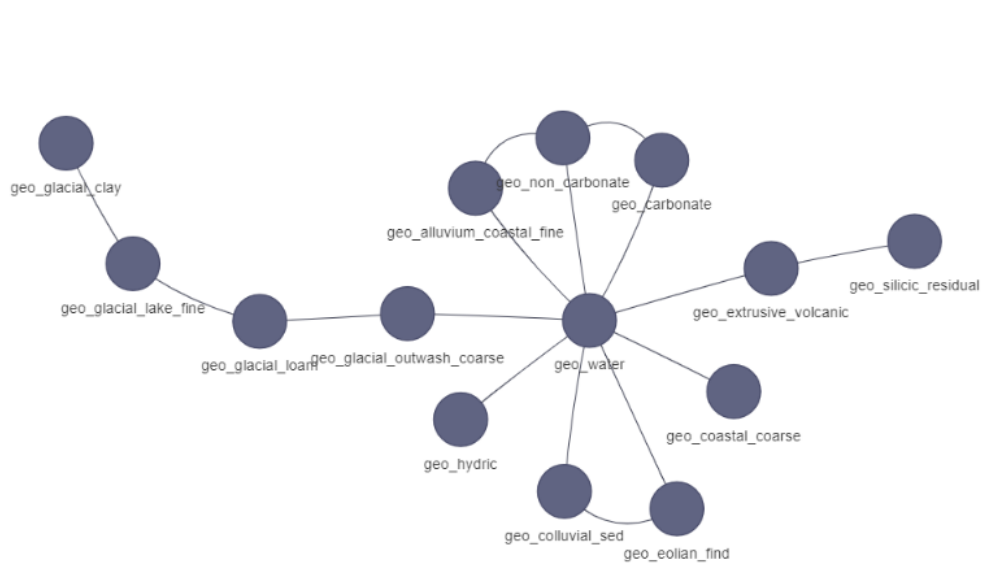


Figure 3: Co-occurrence network for surficial sediments within 1km of NLA lake polygons. Connected nodes were significantly more likely to appear in the same buffer ($p < 0.05$).

Future Work

The influence of sedimentation on lake morphology is not well characterized. Our approach of using sediment categories to relate depth to sedimentation processes was unsuccessful. However, we might have more success including sedimentation in a depth model by determining annual stream flow into lakes.

We also have many terrain metrics to include. Extending basin terrain beyond the lake shoreline only accounts for about $\frac{1}{3}$ of the variance. However, the Heathcote et al. model accounts for nearly $\frac{1}{4}$ of the variance with a single covariate. Other terrain metrics uncorrelated with slope may improve model performance.

References

- ¹ J. Hollister and J. Stachelek, *Lakemorpho: Calculating Lake Morphometry Metrics in R* (F1000Research, 2017).
- ² J. Stachelek, S. Oliver, and F. Masrour, *LAGOSNE: Interface to the Lake Multi-Scaled Geospatial and Temporal Database* (2019).
- ³ J.W. Hollister, W.B. Milstead, and M.A. Urrutia, PLoS ONE **6**, e25764 (2011).
- ⁴ US EPA, *2012 National Lakes Assessment Field Operations Manual* (n.d.).
- ⁵ T.G. Farr, P.A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf, Reviews of Geophysics **45**, RG2004 (2007).
- ⁶ J. Horton, C. San Juan, and D. Stoesser, US Geological Survey Data Series 1052 (2017).
- ⁷ N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, Remote Sensing of Environment **202**, 18 (2017).
- ⁸ R Core Team, *R: A Language and Environment for Statistical Computing* (Vienna, Austria, 2021).
- ⁹ J.A. Veech, Global Ecology and Biogeography **22**, 252 (2013).
- ¹⁰ S. Sobek, Inland Waters **1**, 177 (2011).
- ¹¹ A.J. Heathcote, P.A. del Giorgio, and Y.T. Prairie, Canadian Journal of Fisheries and Aquatic Sciences **72**, 643 (2015).
- ¹² S.K. Oliver, P.A. Soranno, C.E. Fergus, T. Wagner, L.A. Winslow, C.E. Scott, K.E. Webster, J.A. Downing, and E.H. Stanley, Inland Waters **6**, 314 (2016).