Modeling Maximum Depth for all Lakes in the United States

Keenan Ganz

Max Glines

Kevin Rose

# 1 Introduction

A summary of prior models of maximum lake depth is given in Table 4.1.

# 2 Methods

## 2.1 Lake Data Sources

Lake polygons, depth measurements, and other lake attributes were derived from LAGOS-US LOCUS and DEPTH modules1,2. Candidate predictor variables were identified from the four tables included in the LOCUS module. Polygon attributes, including area, perimeter, shoreline development index, maximum distance from shoreline, and oriented bounding box length-width ratio, were calculated within QGIS3.

## 2.2 Terrain Analysis

Terrain metrics were derived from the Shuttle Radar Topography Mission 30m digital elevation model within 200m buffer zones around each lake polygon4. Although other digital elevation models offer finer resolution within the US, SRTM is available worldwide such that the depth model described here may be applied outside of the US.

All terrain analysis was carried out with Google Earth Engine (GEE)5. Predictor variables from the four models summarized in were replicated as closely as possible within the GEE environment. Additional predictors, including buffer elevation, buffer slope, terrain roughness index, horizontal curvature, and kernel-based metrics were included to fully describe the topography of lake buffer zones6. Each metric was calculated for all SRTM pixels within a buffer zone and summarized as the minimum, maximum, mean, and standard deviation.

Kernel-based metrics applied summary statistics to all pixels within a given distance of a central pixel. This captures terrain features at a variety of scales. Terrain relief was calculated as the maximum difference between the central pixel and all other pixels in the kernel. Terrain deviation was calculated as the difference between the central pixel and the mean elevation of all pixels in the kernel, normalized to the standard deviation of all pixels in the kernel.

## 2.3 Spatial Autocorrelation

After Oliver et al. (2016), spatial autocorrelation was assumed at the HUC04 scale. Additional predictors were generated by summarizing terrain metrics, polygon attributes, and lake depth within each HUC04 unit.

## 2.4 Lake Surface Temperature

Lake morphology may be predicted through its relationship with lake heat budget7. When lake waters are well mixed, water temperature is consistent throughout the water column. Under this regime, deep lakes require a larger energy input to illicit a surface temperature change than shallower lakes. Lake surface temperature during turnover may therefore be indicative of maximum lake depth. A pilot study found that EOS ASTER nighttime water surface temperature explained roughly a third of the variance in maximum lake depth.8

Expanding this workflow to all lakes in the United States is complicated by incomplete EOS ASTER nighttime ocverage over the contiguous US, atmospheric effects, cloudy imagery, and regional climate variation. It was hypothesized that lake surface temperature is most strongly correlated with regional climate. Therefore, deviation from a regional median land surface temperature would be indicative of lake heat budget.

Within GEE, daytime 30m Landsat 8 Collection 1 Tier 1 imagery from 2013 to present was used as a source of high-resolution lake surface temperature9. Cloudy pixels were masked with the quality assurance band of each image. Radiance in bands 10 and 11 was converted to water surface temperature according to the procedure given in Vanhellmont (2020)10. First, radiance () was converted to brightness temperature () according to Equation (4.1) and constants listed in Table 4.2.

Brightness temperatures from bands B10 and B11 are then combined and atmospherically corrected according to constants given in Du et al. (2015) (Equation (4.2); Table 4.3). In Equation (4.2), and refer to the average and difference between the two values given in Table 4.2.

Monthly median Landsat surface temperature was calculated at the center of each lake in March-May and September-November to capture periods where lake turnover was most likely.

Monthly median 1km MODIS surface temperature from 2017 was used as representative of local climate11. A focal median filter was applied to the MODIS imagery with a 3-pixel radius circular kernel, effectively increasing the length scale of median surface temperature to 7km. The difference between lake Landsat surface temperature and MODIS regional surface temperature was calculated for each month and included as a separate candidate predictor of maximum lake depth.

## 2.5 Modeling Workflow

### 2.5.1 Model Evaluation

Comparison of existing lake depth models has been impeded by inconsistent source data and small study areas. All models in Table 1 were replicated on the LAGOS-US DEPTH data module. Oliver et al. excluded lakes larger than 1000ha from modeling since depth is known for over 80% of these lakes. A similar proportion (72.5%) of large lakes larger than 1000ha had known maximum depth in the LAGOS-US DEPTH module, so these were excluded from modeling (Figure 4.1). Model performance was evaluated on a stratified sample of 10% of lake depth measurements held out during model training. Evaluation metrics included , root mean square error (RMSE), and mean absolute percent error (MAPE).

### 2.5.2 Random Forest Model

Random forest modeling12,13 was used for its robustness to multicollinearity and ability to capture nonlinear relationships. Feature selection was performed with the R package Boruta to screen out ineffective predictors. The Boruta algorithm proceeds by generating a random forest model for a regression problem and iteratively comparing the importance of a candidate predictor with random noise14. Candidate predictors that outperformed random noise were retained in the random forest model and all others were discarded.

# 3 Results

A summary of model performance statistics is given in Table 4.4.

# 4 Discussion

Table 4.1: Summary of existing models of maximum lake depth, using local topography and regional characteristics.

| Citation | Region | Model | Predictors | Maximum Depth | RMSE |
| --- | --- | --- | --- | --- | --- |
| Oliver et al. (2016) | Northeastern US | Linear mixed effects | Lake area, shoreline development index, buffer slope, watershed to lake area ratio | 0.29 | 7.1m |
| Hollister et al. (2011) | Eastern US | Linear regression | Buffer slope | 0.48-0.67 | 5.09-5.95m |
| Heathcote et al. (2015) | Southern Quebec | Linear regression | Buffer elevation change | 0.52 | 0.245 m |
| Sobek (2011) | Sweden | Linear regression | Lake area, perimeter, buffer slope | 0.36 | Not reported |
| Becker and Daw (2005) | Wisconsin, US | Linear regression | Nighttime water surface temperature | 0.36 | Not reported |

Table 4.2: Constants used in the calculation of brightness temperature from Landsat radiance. Reproduced from Vanhellmont (2020).

| Band | Wavelength (m) | Bandwidth (m) |  | (Wmsrm) | (K) |
| --- | --- | --- | --- | --- | --- |
| B10 | 10.9 | 10.6 - 11.2 | 0.9926 | 774.8853 | 1321.079 |
| B11 | 12.0 | 11.5 - 12.5 | 0.9877 | 480.8883 | 1201.144 |

Table 4.3: Values of constants used to apply atmospheric correction to Landsat brightness temperature from Du et al. (2015)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| -0.41165 | 1.00522 | 0.14543 | -0.27297 | 4.06655 | -6.92512 | -18.27461 | 0.24468 |

Table 4.4: Model validation statistics for the random forest model introduced in this study and selected prior models trained on the LAGOS US DEPTH module. RMSE: root mean square error; MAE: mean absolute error; MAPE: mean absolute percent error; : proportion of variance explained.

| Model | RMSE (m) | MAE (m) | MAPE (%) |  |
| --- | --- | --- | --- | --- |
| Random Forest | 6.732 | 4.415 | 89.509 | 0.417 |
| Oliver et al. (2016) | 7.387 | 4.401 | 68.926 | 0.359 |
| Sobek (2011) | 7.622 | 5.098 | 104.290 | 0.246 |
| Hollister et al. (2011) | 7.804 | 4.762 | 82.438 | 0.247 |
| Heathcote et al. (2015) | 8.629 | 5.126 | 85.201 | 0.119 |

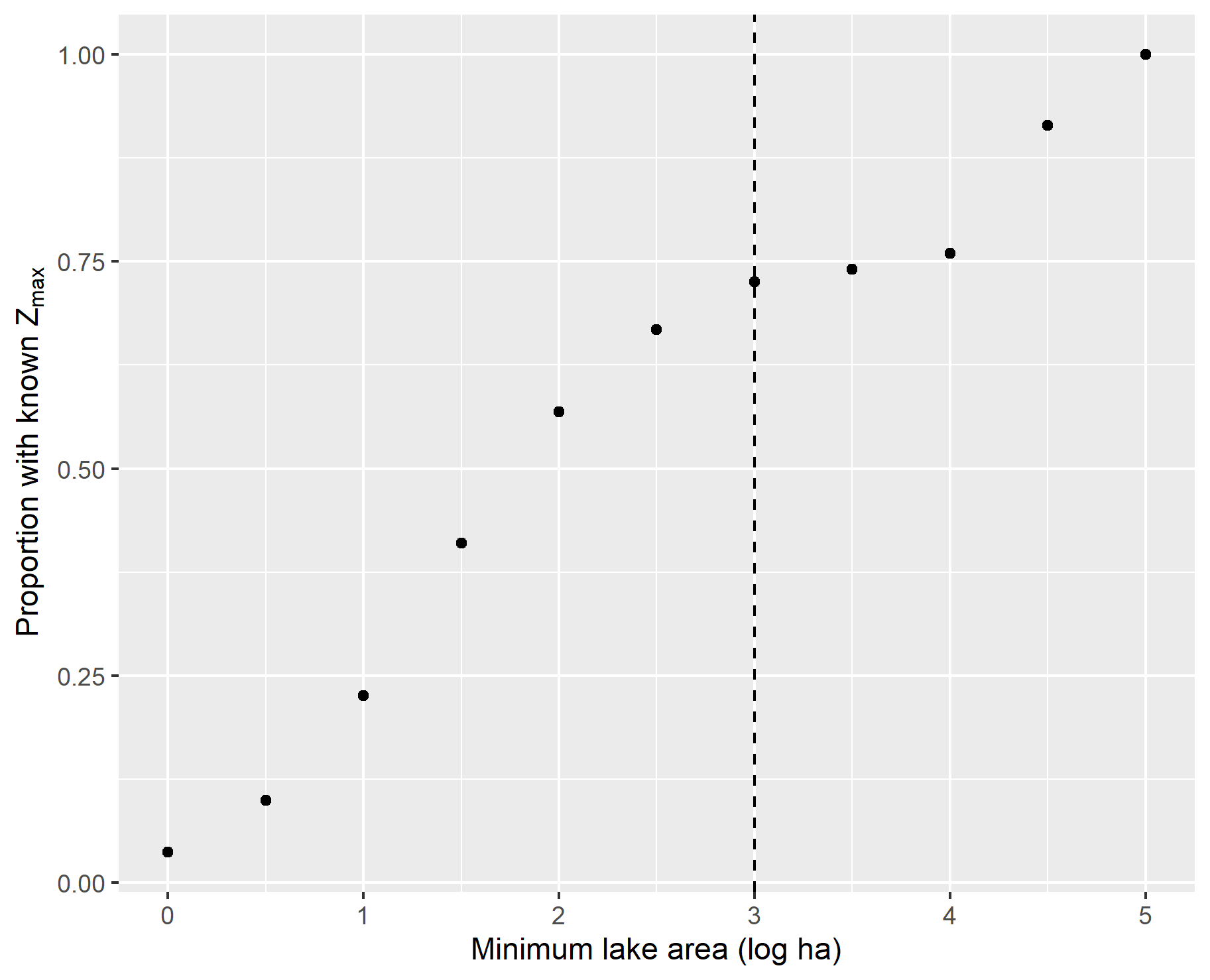


Figure 4.1: Proportion of lakes in LAGOS US with known maximum depth. Vertical dotted line: maximum lake area included in lake depth modeling.

# 5 References

1 K.S. Cheruvelil, P.A. Soranno, I.M. McCullough, K.E. Webster, L.K. Rodriguez, and N.J. Smith, Limnology and Oceanography Letters **6**, 270 (2021).

2 J. Stachelek, L.K. Rodriguez, J. Díaz Vázquez, A. Hawkins, E. Phillips, A. Shoffner, I.M. McCullough, K.B. King, J. Namovich, L.A. Egedy, M. Haite, P.J. Hanly, K.E. Webster, K.S. Cheruvelil, and P.A. Soranno, (2021).

3 Q.D. Team, (2021).

4 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).

5 N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, Remote Sensing of Environment **202**, 18 (2017).

6 J.L. Safanelli, R.R. Poppiel, L.F.C. Ruiz, B.R. Bonfatti, F.A. de O. Mello, R. Rizzo, and J.A.M. Demattê, ISPRS International Journal of Geo-Information **9**, 400 (2020).

7 E.A. Birge, *The Heat Budgets of American and European Lakes* (1915).

8 M.W. Becker and A. Daw, Remote Sensing of Environment **99**, 288 (2005).

9 U.G. Survey, *Landsat 8 Collection 1 Tier 1* (n.d.).

10 Q. Vanhellemont, Remote Sensing of Environment **237**, 111518 (2020).

11 T. Hengl, (2018).

12 L. Breiman, Machine Learning **45**, 5 (2001).

13 M.N. Wright and A. Ziegler, Journal of Statistical Software **77**, 1 (2017).

14 M.B. Kursa and W.R. Rudnicki, Journal of Statistical Software **36**, 1 (2010).

15 J.W. Hollister, W.B. Milstead, and M.A. Urrutia, PLoS ONE **6**, e25764 (2011).

16 S. Sobek, Inland Waters **1**, 177 (2011).

17 A.J. Heathcote, P.A. del Giorgio, and Y.T. Prairie, Canadian Journal of Fisheries and Aquatic Sciences **72**, 643 (2015).

18 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).

19 C. Du, H. Ren, Q. Qin, J. Meng, and S. Zhao, Remote Sensing **7**, 647 (2015).