Direct Morphological Modeling of NLA Lakes, Revisited

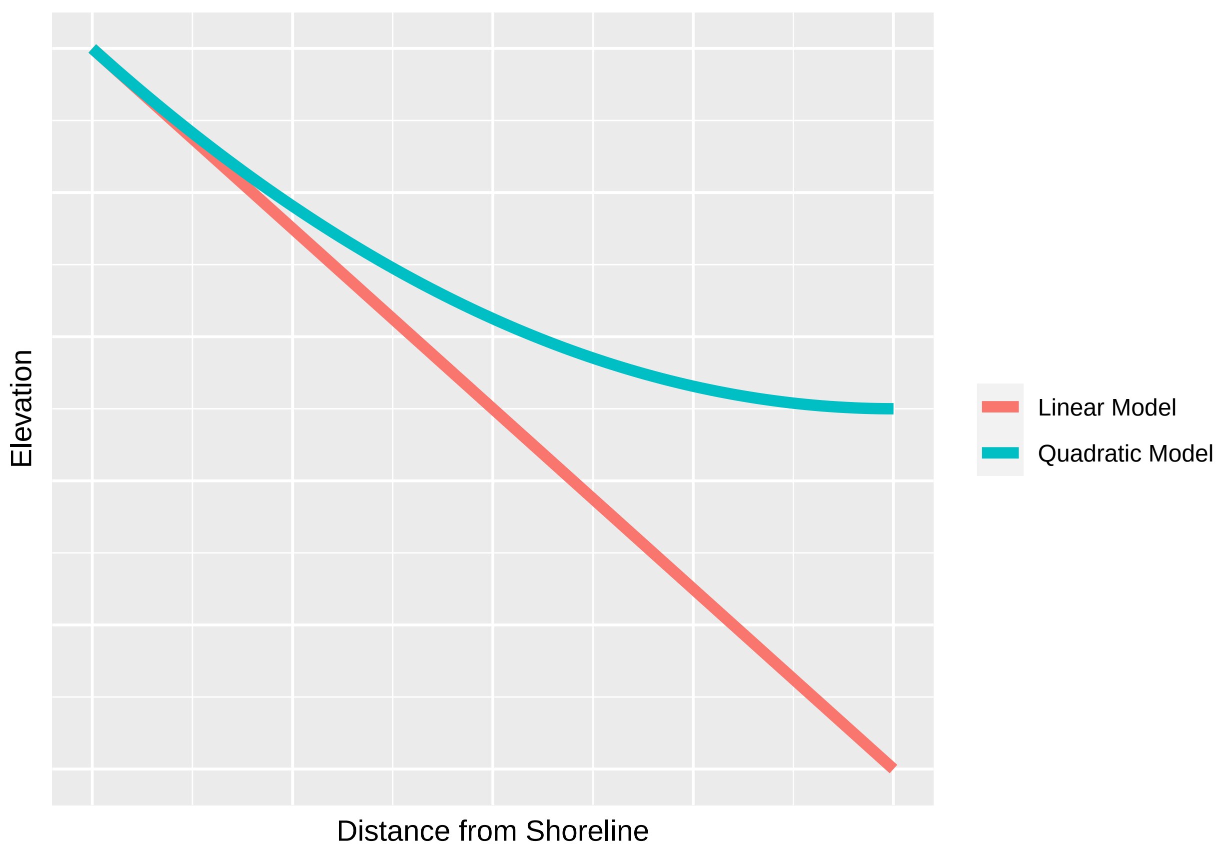
Keenan Ganz

Introduction

In `02_nla_morpho`, we reproduced the lake depth modeling technique in Hollister et al[1]. In short, this method linearly extends the surrounding terrain to estimate lake depth. Median slope in a buffer region is multiplied by the maximum distance from shoreline (Eq. 1)

$$\hat{Z} = S \times D_{max} \tag{1}$$

The cited study considered a suite of NLA lakes in the western United States with a dynamic buffer region. When this model is applied to all unique NLA lakes from 2007 and 2012, we found that it generally overestimates lake depth. Another function, such as a quadratic curve, may be closer to reality. If we consider a profile from the shoreline to the point of maximum depth, the model result is as follows.



The linear model is easily calculated from the median slope of the surrounding landscape. To calculate the quadratic model, we can consider the second-order Taylor approximation of a differentiable polynomial (Eq.

2).

$$\hat{Z} = S_1 \times D_{max} + \frac{S_2}{2} \times D_{max}^2 \tag{2}$$

Here, S_1 is the median terrain slope in the buffer region and S_2 is the median profile curvature in the buffer region. Profile curvature (also known as vertical curvature) is equivalent to the second derivative of elevation, calculated in the direction of maximum slope. A detailed discussion of morphology variables is provided in Florinsky[2].

If the quadratic model is more effective than the linear model, we would expect that for most lakes, $S_1 < 0$ and $S_2 > 0$. We will test these assumptions and then determine if a quadratic model is a more effective estimator of maximum lake depth.

Methods

Data Sources

Polygons and maximum depth information from the EPA National Lakes Assessment (NLA) were downloaded from the US EPA website (https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys). Buffer strips 200m wide were computed in QGIS and uploaded to Google Earth Engine (GEE)[3]. For elevation data, we used the elevation product from the Shuttle Radar Topography Mission at 30m cell size[4].

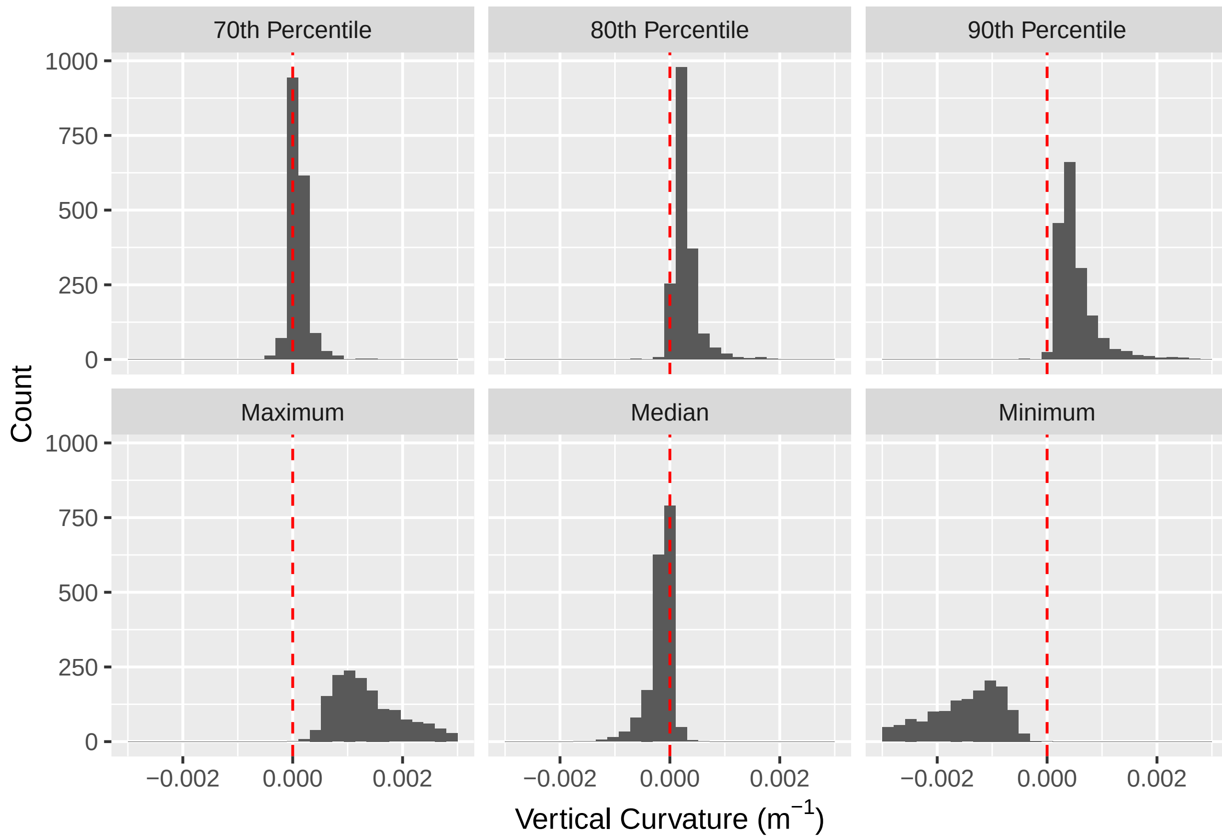
Spatial Analysis

Some buffer strips were especially complex and had to be simplified for processing in GEE. The maximum error in simplification was 30m. In GEE, slope was calculated from a built-in function. To calculate vertical curvature, we used the Terrain Analysis in GEE (TAGEE) package[5]. The raw SRTM data was smoothed with a Gaussian filter with radius equal to three pixels and $\sigma = 2$ before processing by TAGEE. Within each buffer region, the minimum, maximum, median, and selected percentiles were calculated for vertical curvature values within the buffer. Statistics were downloaded as a CSV and analyzed further in R[6].

Results

Exploratory Data Analysis

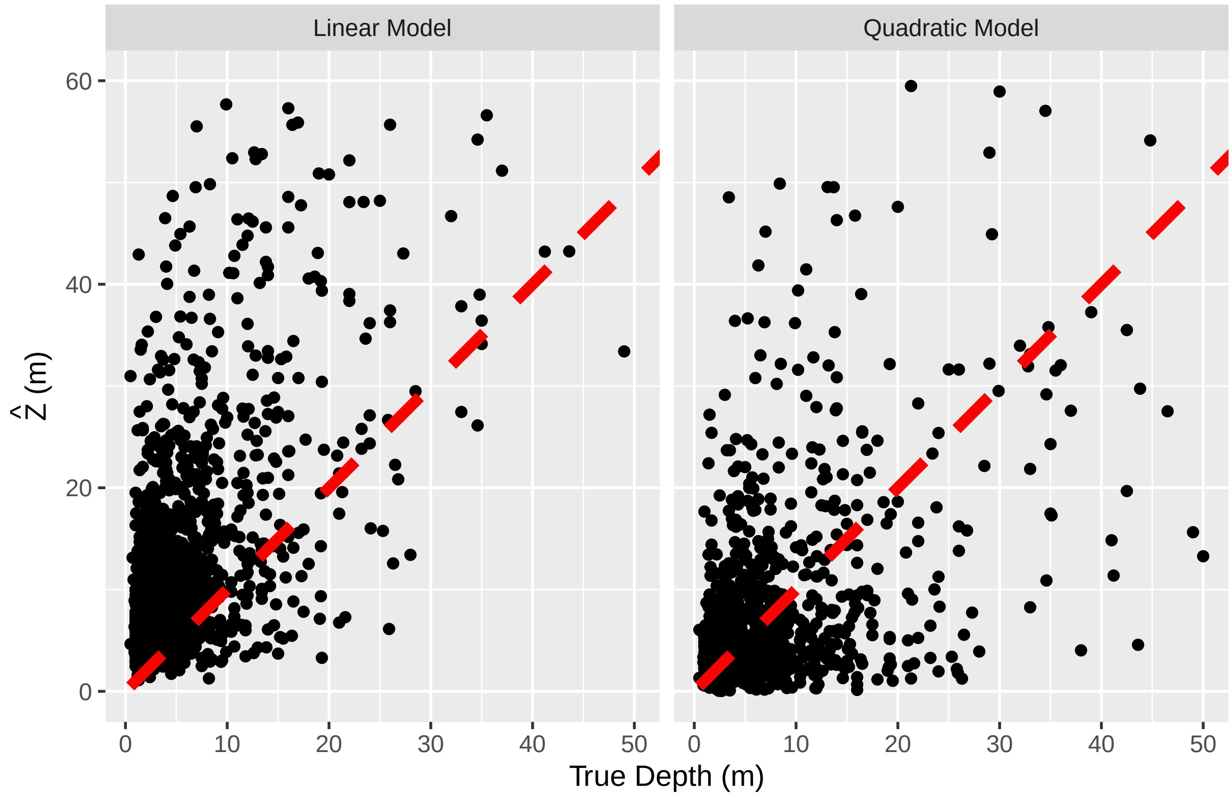
We begin by determining if our assumption that the majority of lakes have positive curvature.



Statistic	Percent Greater than Zero
70th Percentile	80.66
80th Percentile	97.86
90th Percentile	99.89
Maximum	100.00
Median	15.57
Minimum	0.00

The 50th percentile vertical curvature in the majority of buffer regions is negative, but the 70th percentile vertical curvature is positive. Given that we know the linear model overestimates lake depth, the quadratic model would be less accurate if we use median vertical curvature. Instead, we will use the 90th percentile.

Depth Modeling



Model	R^2	RMSE
Linear	0.38	21.66
Quadratic	0.25	10.61

We can also try a multiple linear modeling approach where the linear and quadratic terms are predictors.

```
##
## Call:
## lm(formula = maxdepth ~ dist_pole:slope200_median + dist_pole_squared:p90,
##     data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.800  -3.732  -1.894   1.789  40.122
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  4.667e+00  2.039e-01   22.88   <2e-16 ***
## dist_pole:slope200_median    1.269e-01  4.172e-03   30.41   <2e-16 ***
## dist_pole_squared:p90       -7.572e-04  5.929e-05  -12.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.192 on 1776 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3435, Adjusted R-squared:  0.3427
## F-statistic: 464.6 on 2 and 1776 DF,  p-value: < 2.2e-16
```

Finally, we can use a general additive model (GAM) which uses the linear and quadratic terms as predictors. First, we use only the linear term.

```
##
## Call:
## lm(formula = maxdepth ~ dist_pole:slope200_median, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.655  -3.995  -2.076   1.752  40.403
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               5.214636   0.208274   25.04   <2e-16 ***
## dist_pole:slope200_median 0.101256   0.003822   26.50   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.513 on 1777 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2832, Adjusted R-squared:  0.2828
## F-statistic:   702 on 1 and 1777 DF,  p-value: < 2.2e-16
```

The deviance explained is similar to other linear models we have seen.

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## maxdepth ~ s(linear_term) + s(quad_term)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.0744     0.1587   50.87   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                  edf Ref.df      F p-value
## s(linear_term) 5.707  6.757 114.17  <2e-16 ***
## s(quad_term)   7.918  8.595  18.81  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.431   Deviance explained = 43.5%
## GCV =  45.19  Scale est. = 44.819    n = 1779
```

When we add the quadratic term, the Deviance explained increases by about 50%.

Conclusions

Including profile curvature as a quadratic term in the linear depth modeling equation improves model accuracy by preventing overestimation in the linear term. In order to directly compare the quadratic model to the linear model, we must completely reproduce the spatial workflow presented in Hollister et al.

References

[1] J.W. Hollister, W.B. Milstead, and M.A. Urrutia, PLoS ONE **6**, e25764 (2011).
[2] I.V. Florinsky, *Digital Terrain Analysis in Soil Science and Geology*, 1st ed (Elsevier/Academic Press, Amsterdam ; Boston, 2012).
[3] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, Remote Sensing of Environment **202**, 18 (2017).
[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] 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).
[6] R Core Team, *R: A Language and Environment for Statistical Computing* (Vienna, Austria, 2021).