Geologic Context of NLA Lakes, Revisited

Keenan Ganz

Introduction

In `01_nla_geo.pdf`, we looked at a dataset of surficial geology and determined whether the parent sediment types had any predictive power on lake depth. We found that the predictive power was overall quite low ($R^2 \approx 0.1$), but we did see interesting cooccurrence of sediments according to a physical process. For example, we saw that glacial sediments sorted themselves by grain size.

The source dataset for that notebook was part of a larger publication on landforms for climate adaptation planning[1]. That raster was in turn derived from surficial material mapping undertaken by the US Geologic Survey[2]. Although surficial materials are ecologically important, they may not have as large an impact on terrain morphology as underlying bedrock. This notebook will revisit the analysis from the first notebook with the State Geologic Map Compilation[3].

Methods

SGMC Characteristics

The SGMC is a compilation of state geologic maps. As such, the SGMC is not standardized across state boundaries. There are five nested levels of lithology categories. We only consider the primary and secondary categories, which are listed below.

Primary Category	Secondary Category
Sedimentary	Carbonate
	Chemical
	Clastic
	Coal
Igneous	Plutonic
	Volcanic
	Hypabyssal
Metamorphic	Metaigneous
	Gneiss
	Granoblastic
	Schist
	Metasedimentary
	Amphibolite
	Migmatite
	Hydrothermally-altered
	Granulite
Tectonite	Melange
	Mylonite
Unconsolidated	Fine-detrital
	Coarse-detrital
	Peat
	Marl

The SGMC is desirable because of its nested design. The schema is also designed from a geologic perspective.

Spatial Analysis

As in the previous notebook, we begin with sampled lake polygons from the 2007 and 2012 EPA National Lakes Assessment[4]. Duplicates across the two sampling campaigns were removed, and the polygon was buffered by 1km. Both of these datasets were uploaded to Google Earth Engine[5]. SGMC polygons were rasterized to 50m cell size and the relative areal cover of each lithology category was computed.

Cooccurrence Analysis

To determine if certain lithology categories are more likely to cooccur in the 1km buffer region around each lake, we used the species cooccurrence model implemented in the R package `cooccur`[6].

Results

Exploratory Data Analysis

Although constructed in a different way, the SGMC data should hav esimilar characteristics as the previous surficial geology data. Like before, we expect that the majority of lakes are associated with one major geology type.
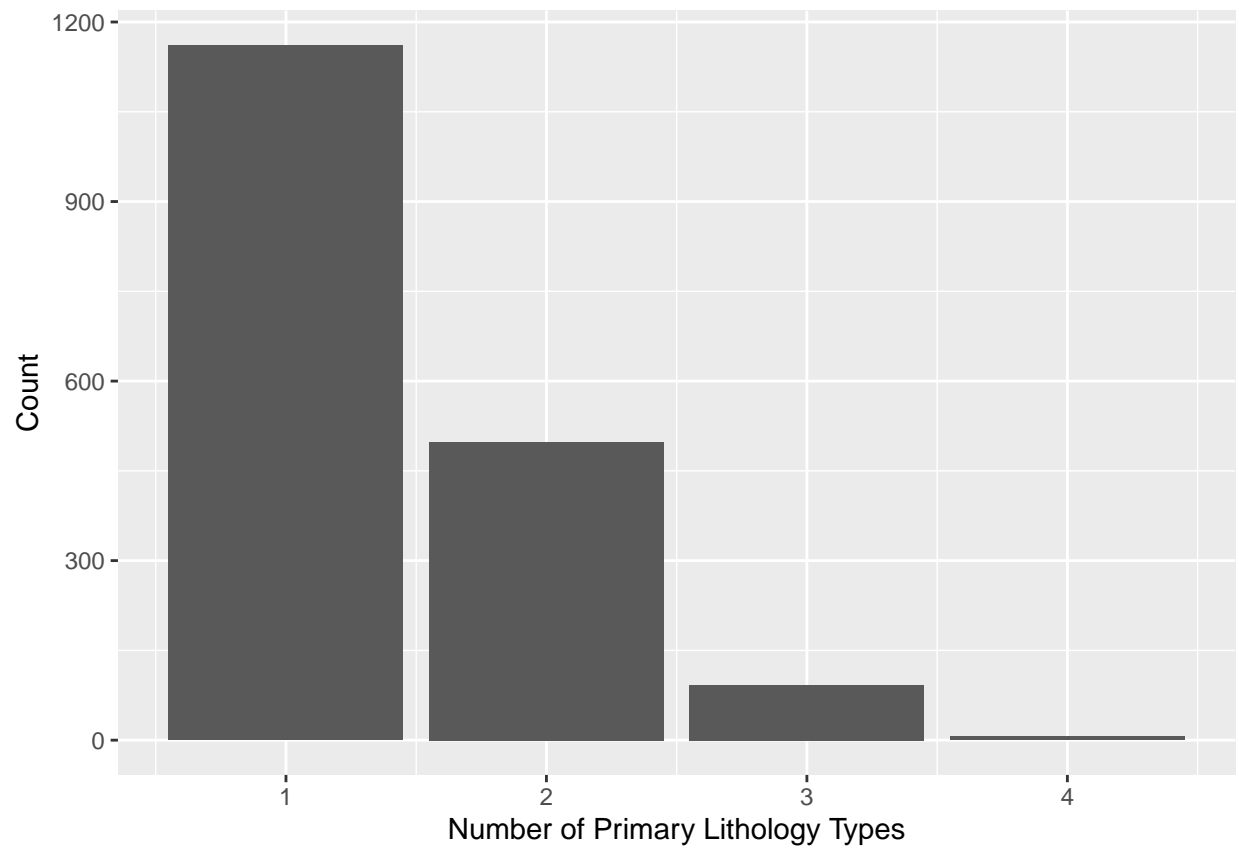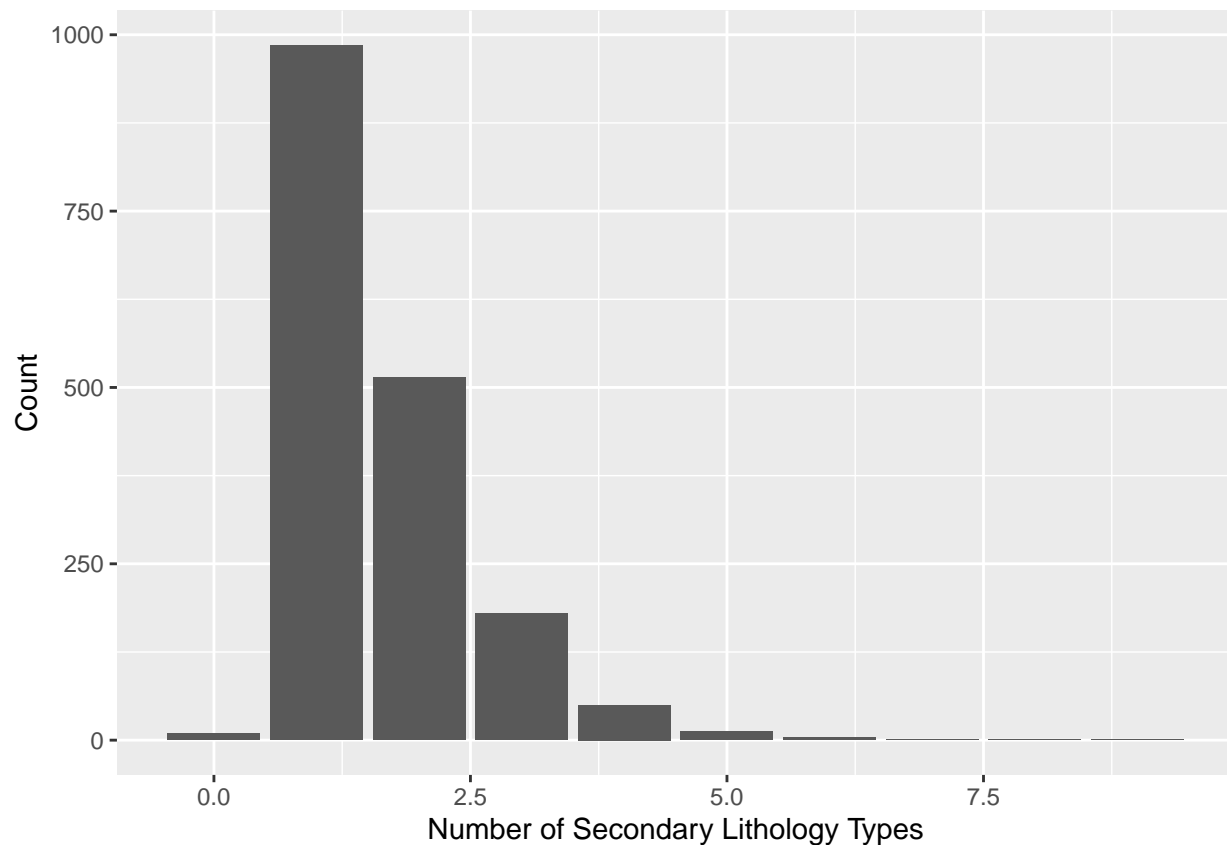
As before, most lakes have 1-2 lithology types. Now, let's see if any category is associated with lake depth. For this, we assign lakes to lithology types by computing the category with the largest areal cover in each polygon. Only two lakes had tectonite as their dominant lithology, so this category was discarded for this section. Likewise, there were few lakes with peat, migmatite, or granulite as the dominant secondary lithology, so these were discarded as well.

Dominant Primary Lithology
- Igneous
- Metamorphic
- Sedimentary
- Unconsolidated

Log Maximum Depth
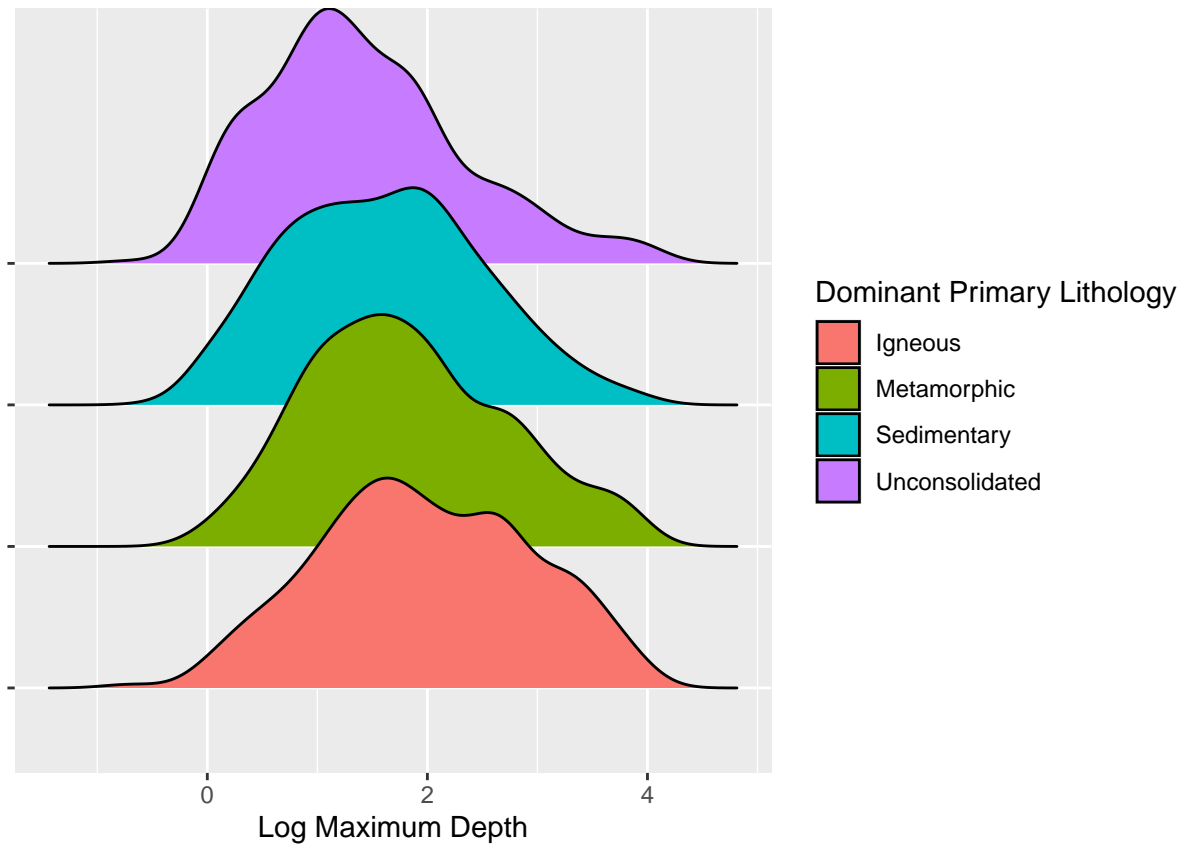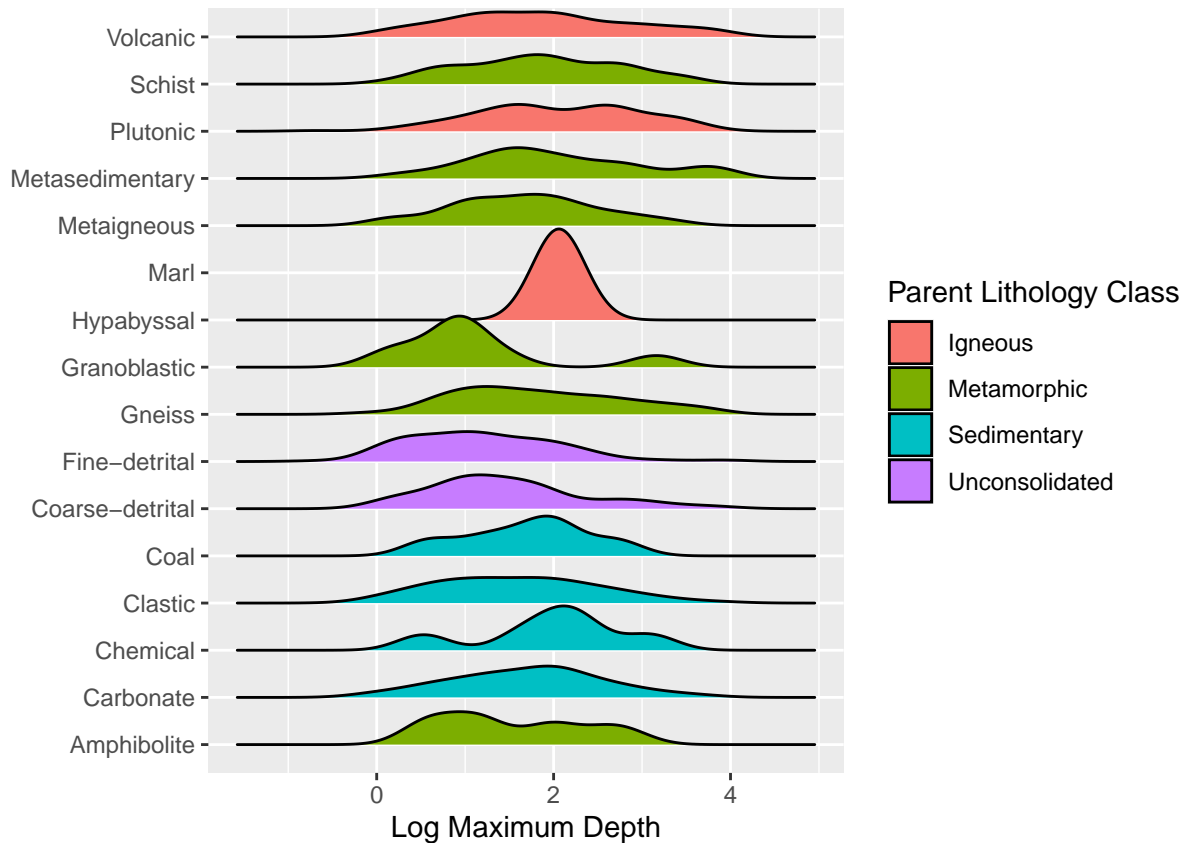
With a log transformation, the distributions appear normal. However, a Shapiro-Wilk non-normality test was significant for each category. Instead, we use the Kruskal-Wallis test as an alternative to a one-way ANOVA. Bear in mind that there is some spatial autocorrelation in the data for lakes that are close together.

The test has $p =$ 0. This is significant, so this may indicate that lithology plays a role in determining maximum lake depth. However, the distributions in the above density ridges plot appear quite similar. So, the predictive power is probably poor.
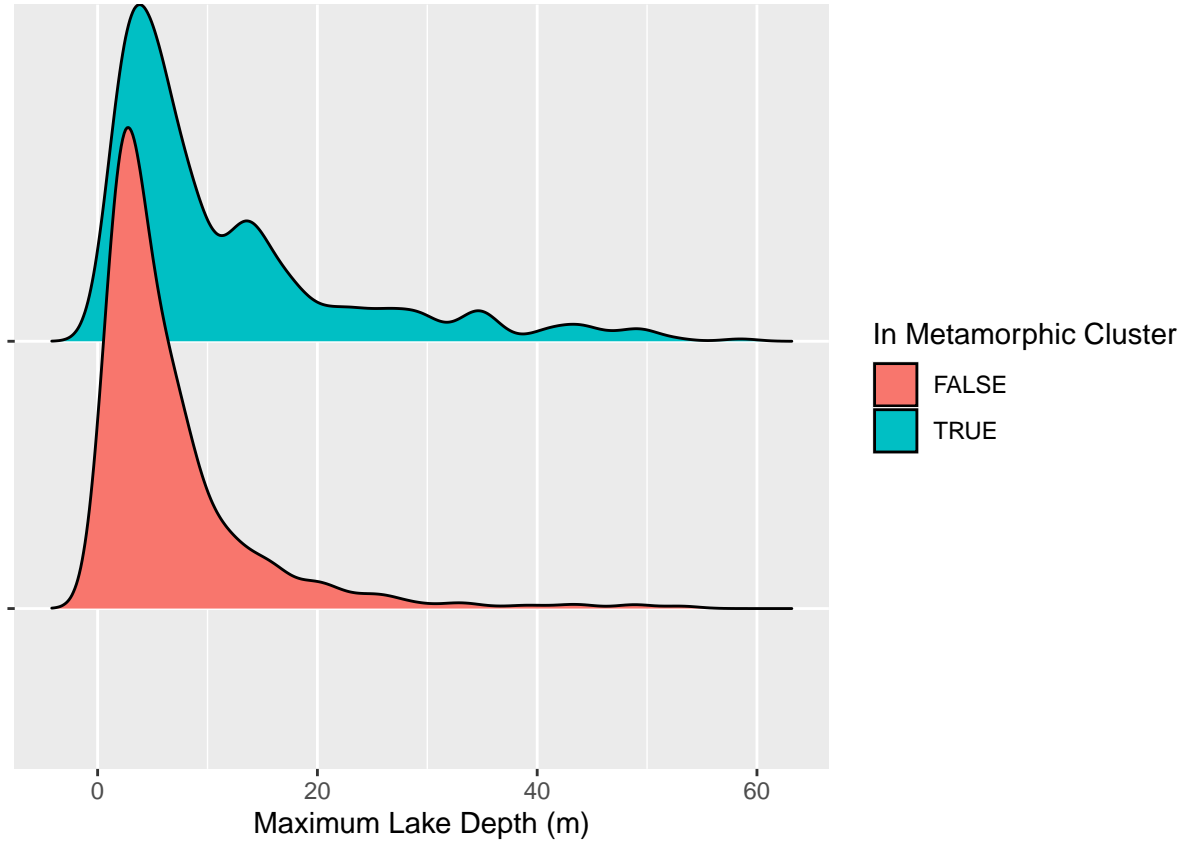
Cooccurrence Analysis

If cooccurring sediments align into clusters, we may be able to split lakes by their environment.

```
##    |                                                              |
```

Secondary Lithology Co−occurrence Matrx
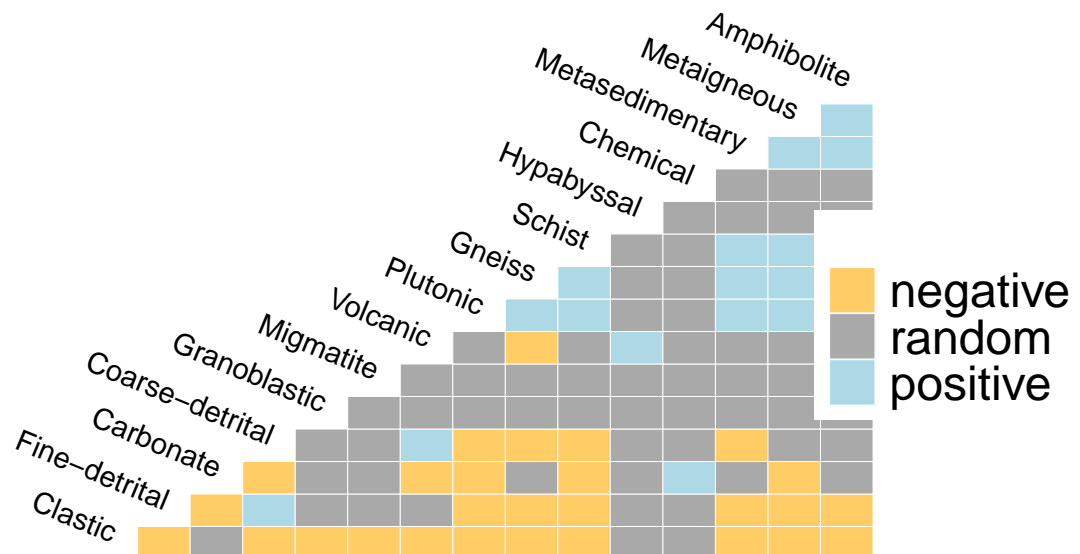
Several metamorphic secondary lithologies cluster together while the majority of secondary lithologies are disconnected. Does the single cluster have a different depth?

Lakes that are in the metamorphic cluster have a longer upper tail than those lakes outside of the cluster. This difference is significant according to a Kruskal-Wallis rank sum test ($p < 10^{-3}$).

Random Forest Modeling

Finally, we consider the use of random forest in predicting maximum lake depth. Since the lithology types are nested, we used only the secondary categories to form this model.
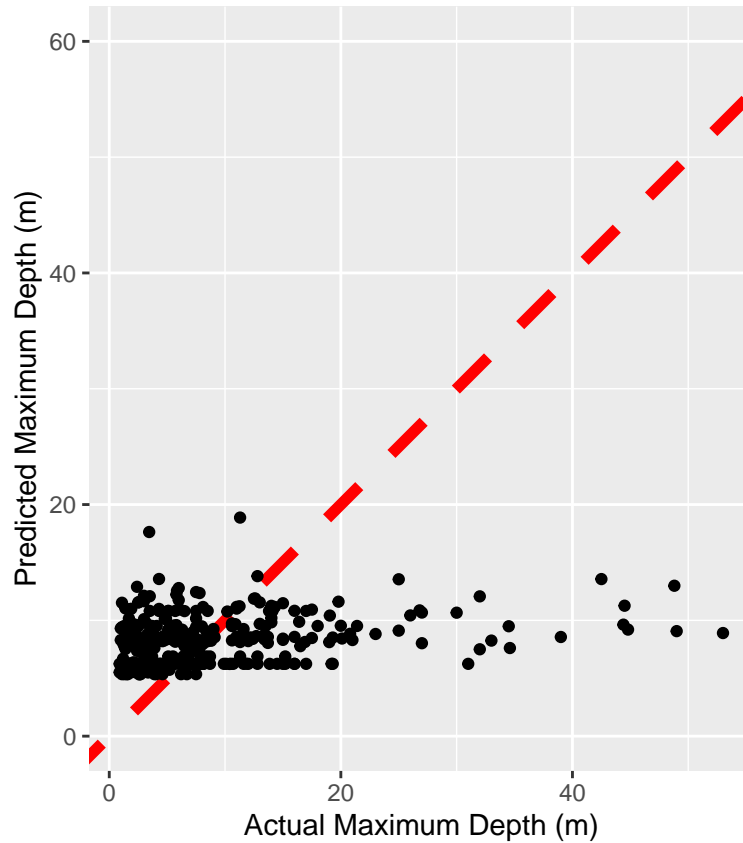
This model had an R^2 of 0.08 and RMSE of 8.46m.

Conclusions

As with the previous geologic dataset, we uncovered minor differences in lake depth along lithologic categories. Some cooccurrence structures were identified, but these did not give any more predictive power than the previous dataset.

References

[1] D.M. Theobald, D. Harrison-Atlas, W.B. Monahan, and C.M. Albano, PLOS ONE **10**, e0143619 (2015).

[2] D. Soller, M. Reheis, C. Garrity, and D.V. Sistine, US Geological Survey Data Series 425 (2009).

[3] J. Horton, C. San Juan, and D. Stoeser, US Geological Survey Data Series 1052 (2017).

[4] U. EPA, *2012 National Lakes Assessment Field Operations Manual* (n.d.).

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

[6] J.A. Veech, Global Ecology and Biogeography **22**, 252 (2013).