Geologic Context of NLA Lakes

Keenan Ganz

```
knitr::opts_knit$set(root.dir = normalizePath(".."))
knitr::opts_chunk$set(
    echo = FALSE,
    message = FALSE,
    warning = FALSE
)
```

Introduction

Lake morphology, including depth and volume, is possibly influenced by the surficial geology of the surrounding area. Distinct lakes from the 2007 and 2012 EPA National Lakes Assessment were identified and 1km buffer zones around each lake were produced. The relative areal cover of 20 categories of parent soil material was calculated in each buffer region[1].

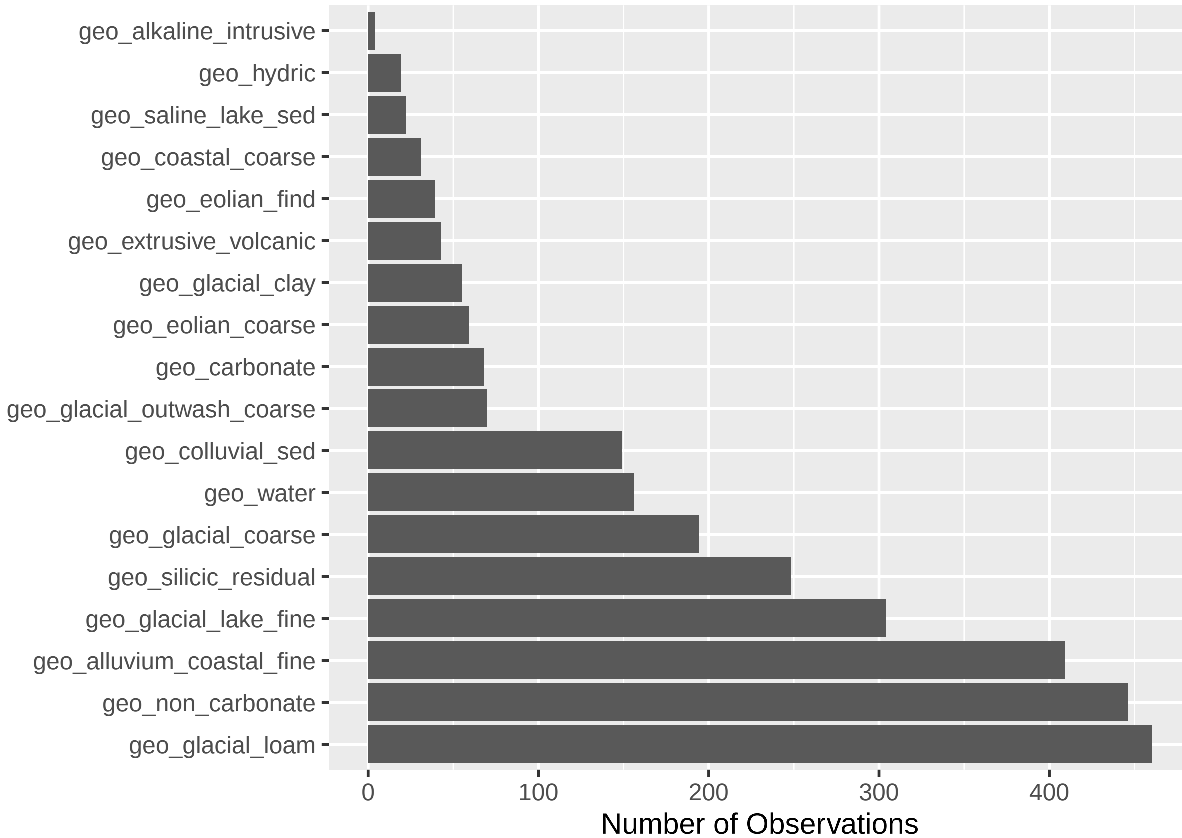
The resulting dataset is high-dimensional. Principal Component Analysis (PCA) decomposes high-dimensional datasets into a low-dimensional dataset or orthogonal factors. Clustered observations in principal component space may be indicative of the geologic contexts in which lakes form.

The categories of parent soil material are described in Table 1.

Table 1: Lithology classes and their 3-letter reference codes used in the text.

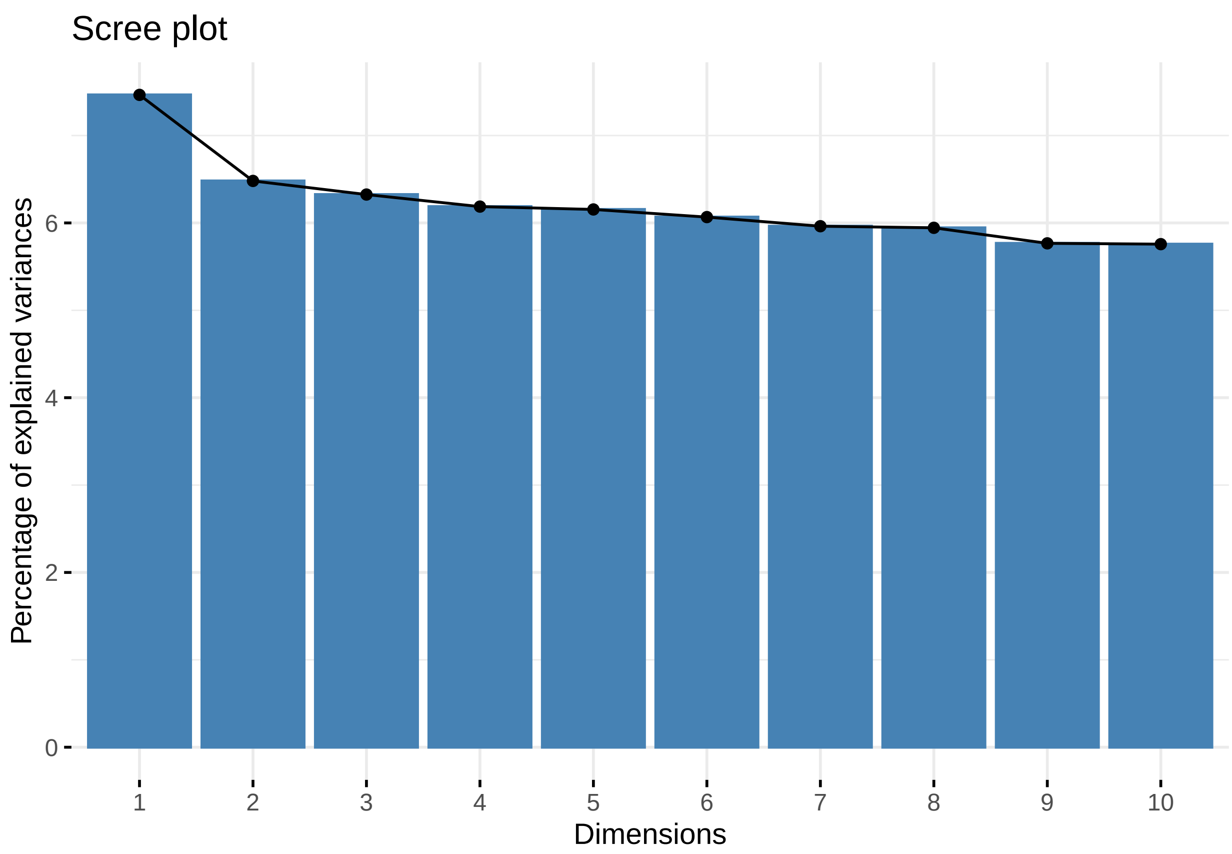
Number	Code	Description
0	WAT	Water
1	CAR	Carbonate
3	NCB	Non-carbonate
4	AKI	Alkaline intrusive
5	SIR	Silicic residual
7	EXV	Extrusive volcanic
8	COS	Colluvial sediment
9	GTC	Glacial till clay
10	GTL	Glacial till loam
11	GTC	Glacial till coarse
13	GLS	Glacial lake sediment fine
14	GOC	Glacial outwash coarse
15	HYD	Hydric
16	ESC	Eolian sediment coarse
17	ESF	Eolian sediment fine
18	SLS	Saline lake sediment
19	ACS	Alluvium and coastal sediment fine
20	CSC	Coastal sediment coarse

First, let's consider the number of times each sediment type is observed.

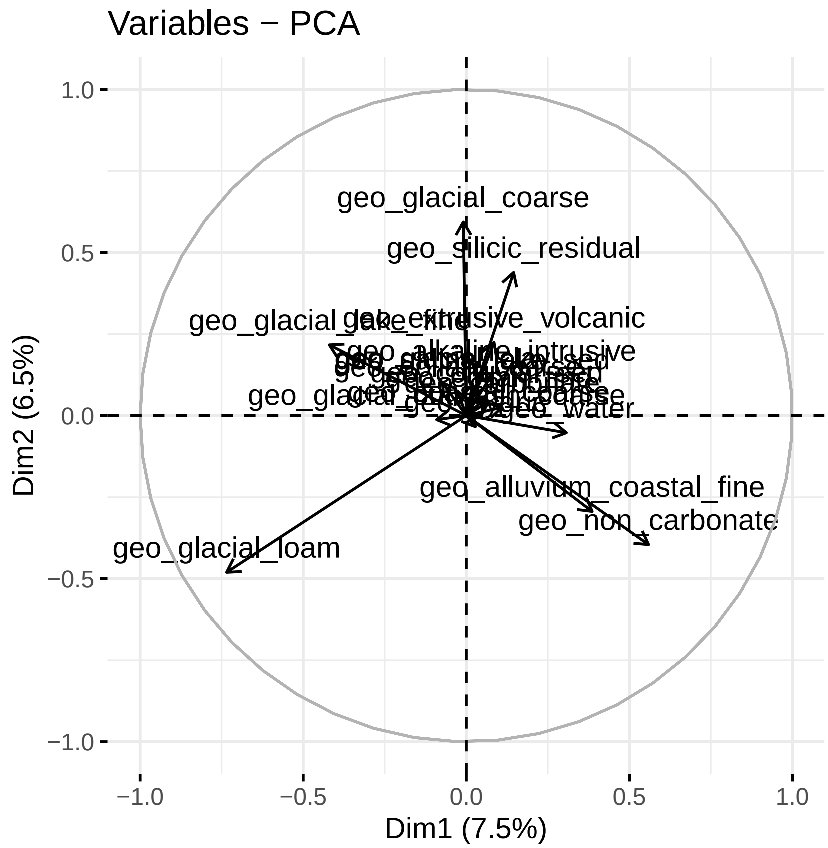


Dimensionality Reduction

We begin with a scree plot for a PCA.

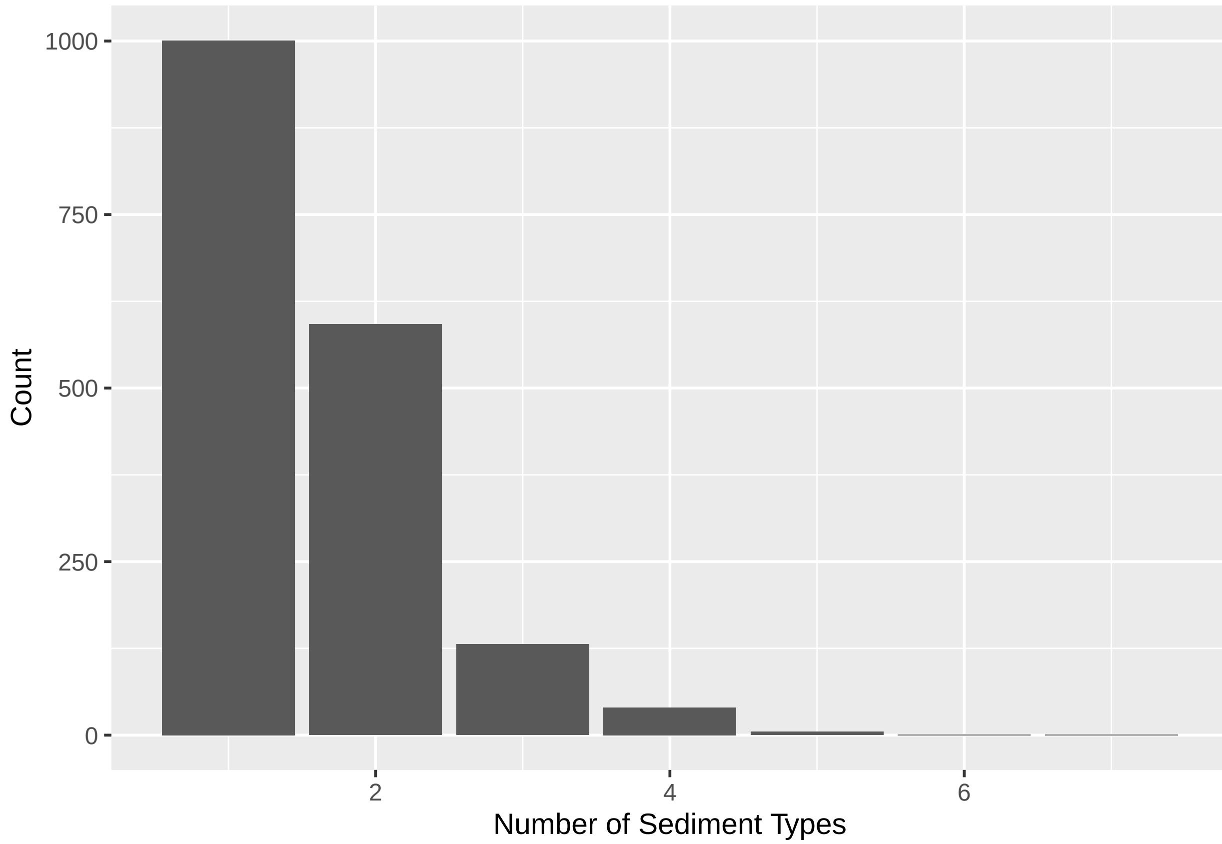


The elbow in the scree plot occurs at two principal components, so only these will be retained. Only a small proportion of the variance can be explained by this decomposition.



Few geology types load onto the components since a low proportion of variance is explained. Only the glacial variables show large loadings while the other variables cluster near the origin.

The majority of observations are "all or nothing." That is, the geologic context of a lake is usually only one sediment type. Consider the number of sediment types present at each lake.

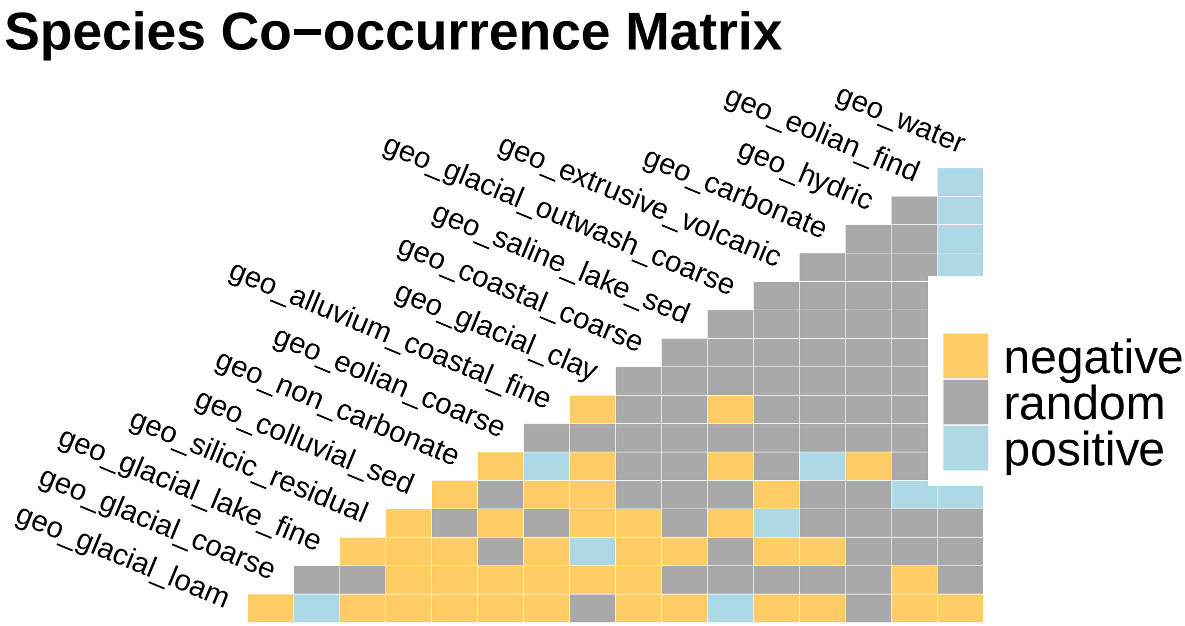


Dimensionality reduction will not work well on this data because the matrix is sparse. A better analysis technique may be co-occurrence analysis.

Co-occurrence Analysis

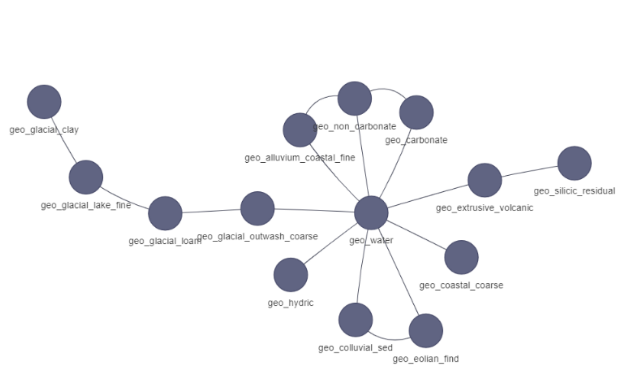
A probabilistic co-occurrence model was applied to a presence-absence matrix of sediment types[2]. Sediment interactions are categorized as either negatively co-occurring, positively co-occurring, or random (i.e. insignificant relationship). The interaction matrix is below.

```
##    |                                                                          |
```



Most of the interactions are either negative or insignificant. Co-occurring sediment groups include certain glacial sediments as well as water and coastal sediments. These groups may be better visualized in a network.

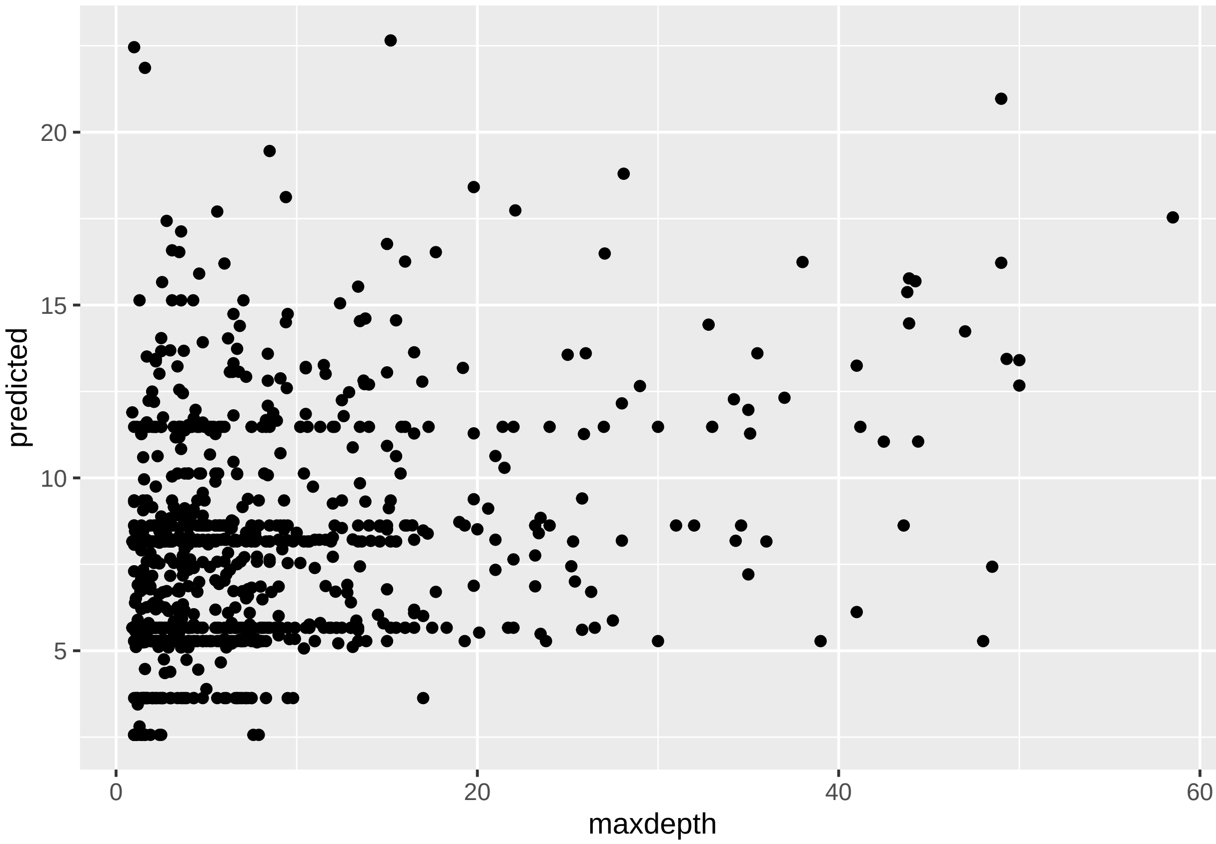
In the below figure, solid lines indicate positive co-occurrences while dashed lines indicate negative co-occurrences.



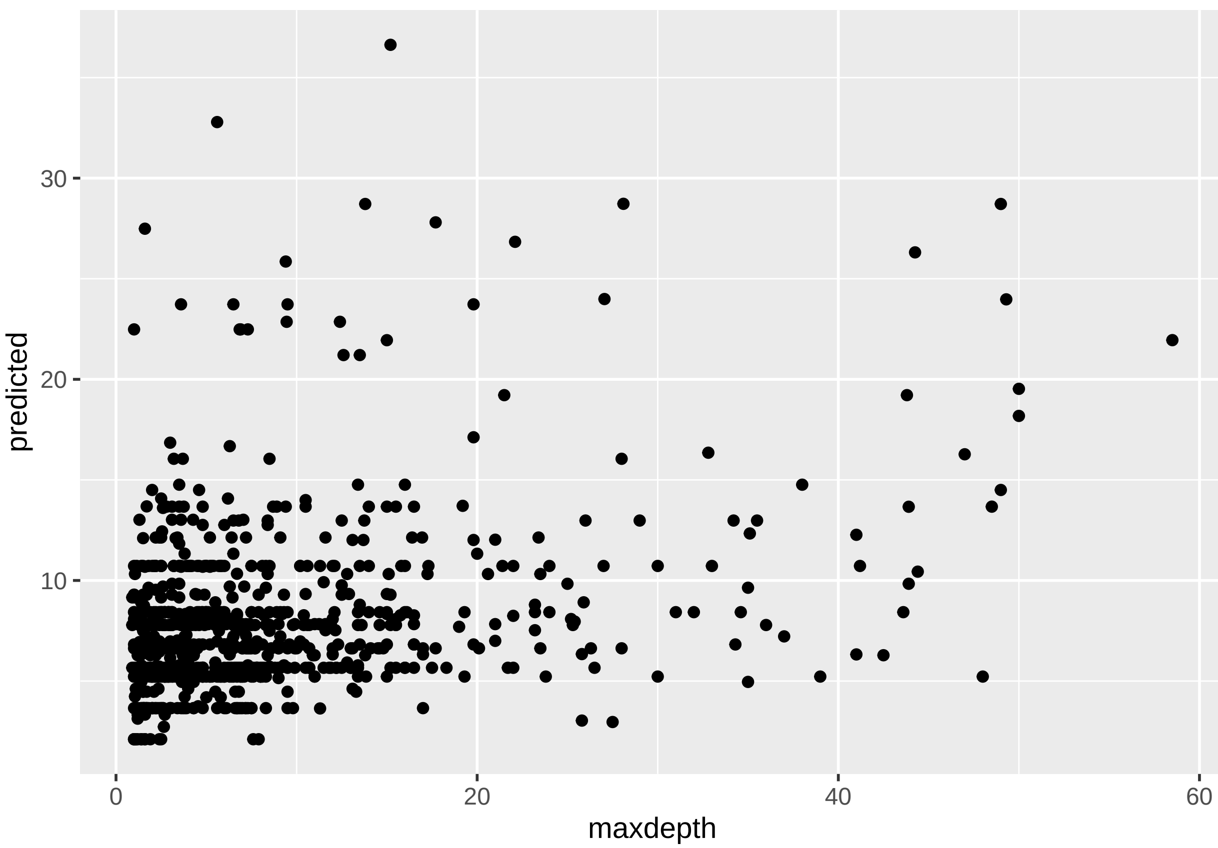
Water is the most well-connected sediment type. This is unsurprising, since all of these data come from lakes. Where one lake is, other water bodies are also probably nearby. The rarest sediment types are isolated nodes. Finally, the glacial nodes have organized themselves according to grain size, with the exception of the coarsest grains.

Random Forest Model

We have identified which sediments tend to appear in the same lake buffer. But, does this tell us anything about the lake morphology? With many variables that likely have a nonlinear relationship with depth, random forest is an appropriate modeling approach.



Does the RMSE improve if we use the categorical form of the data?



The results are comparable between the categorical and continuous models, which had maximum depth RMSE of 8.36m and 8.37m respectively.

Conclusions

Although some cooccurrence relationships were observed between sediment types, sediments are not a particularly useful predictor of maximum lake depth. By comparison with other studies, the predictive power of surrounding sediment is slightly less than that of median buffer slope and lake area[3]

References

[1] D.M. Theobald, D. Harrison-Atlas, W.B. Monahan, and C.M. Albano, PLOS ONE **10**, e0143619 (2015).
[2] J.A. Veech, Global Ecology and Biogeography **22**, 252 (2013).
[3] S. Sobek, Inland Waters **1**, 177 (2011).