Analysis for Hollister, Kreakie, Milstead

Jeff Hollister, Betty Kreakie, Bryan Milstead

Monday, April 28, 2014

Main Question: Given that we have a unique combination of novel data, increased computational facilities, and an expanded geographic scope, what can we add to the classic understanding of the processes that are predictive of chlorophyll *a* and trophic state in lakes.

Initital variable selection (Jeff)

1.mean temp 2.DD45 3.Select scale(s) - local - 300m and regional - 3000m *ANALYSIS*- as a check look at centroid distance matrix 4.Turbidity, not secchi - because turb is an instrument measure and not limited by depth of lake. If need be Turb could be converted to and estimated secchi.

Analysis 1. Compare classic linear models to random forest models?

1.Chl a ~ TP (converted to TS) 2.Chl a ~ TN (converted to TS) 3.Chl a ~ TN + TP (converted to TS) 4.Chl a ~ Linear Mod with varSelRF variables (converted to TS) 5.Chl a TS ~ RF(varSelRF variables) - single randomForest run 6.Chl a TS ~ Random Forest Consensus - single randomForest run

ANALYSIS - varSelRF (Bryan plus play with varSelRFBoot) - run linear models (1-4) (Bryan) - run RF (Betty) - randomForest All Variables - randomForest varSelRF Variables - output error matrices (Bryan, Betty and Jeff) - overall accuracy - Kappa

Question 2. Are *in situ* nutrient and water quality data required to make predicitions of trophic state or can GIS derived information on lake and land use/land cover also provide quality predictions?

Predict for these classes 1. NLA Chl a trophic state classes 2. High/Low trophic state classes

Question3. Can data-mining approaches identify new, interesting variables to use for the prediction of trophic state?

Here's the analysis part:

Initial Variable Culling:

This set of code creates namesets needed to cull out the initial set of variables

Chl *a* Trophic Status ~ All Variables + Landscape Percent

Chl *a* Trophic Status ~ GIS Only Variables +Landscape Percent