A State-Scale Investigation of Statistical Relationships Between Lake Water Quality and Multiple Lake Watershed Characteristics

Brendan Hennessey: Department of Geography, University of Vermont. bthennes@uvm.edu
Dr. Beverley Wemple: Department of Geography, University of Vermont. bwemple@uvm.edu

Abstract

The water quality of freshwater lakes is important both for the ecosystems they support, as well as for human use and recreation. Watershed characteristics and land use are a facet of lake health often overlooked in favor of more local processes. I hypothesized that at the state-level scale, significant statistical relationships exist between watershed metrics (including percent agricultural land, percent wetland, road density, mean elevation, mean slope, and lake-to-basin size ratio) and various measures of lake health, including TN, TP, and a composite Water Quality score developed by the Vermont Agency of Natural Resources.

Materials and Methods

The mean TP, TN, and Chloride data were collected as part of the Vermont Agency of Natural Resources annual lake sampling program. The values are derived from top and bottom measurements collected from the years 2011 to 2014, although few lakes were sampled more than twice. A "Water Quality Score" was given to Vermont lakes based on TP, Chloride, and Secchi Depth Transparency trends observed in available sampled data. The Lake outlines and watersheds were provided by the Vermont ANR, through Dan Homeier.

Example Metric Extraction Processes:
(Analysis completed in ESRI ArcGIS)

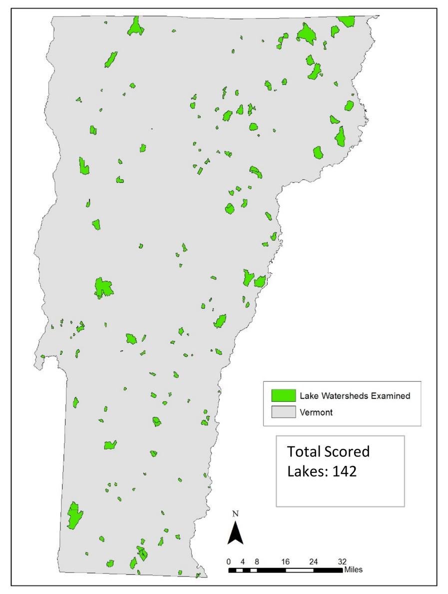
Lake-to-basin area: Spatial Join → join data table → calculate new field as lake area/basin area

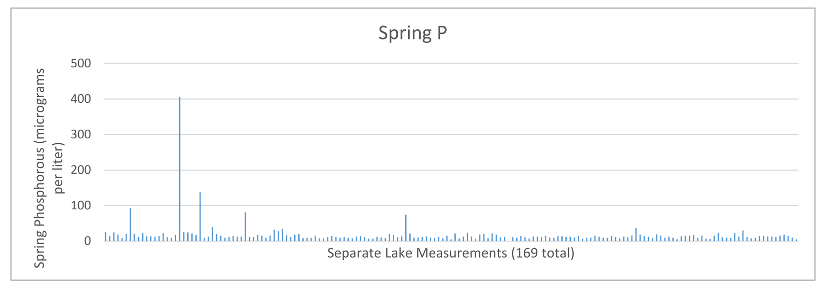
Roads: Intersect road layer (E911) and watershed layer (Dan Homeier) → select field LakeID → Summarize (sum) → join to watershed layer → new field, calculate as "=Shape_length"→ Switch selection, fill rest with 0

Land Cover: Extract by Mask (with NLCD 2011 as input and Watershed layer as mask)→ Extract by Attributes (SQL query for wetland/ag) → Zonal stats as table → Join to watershed layer → calculate new field as area for LC_type (count*900) → Calculate percent as (previous field/total area *100)

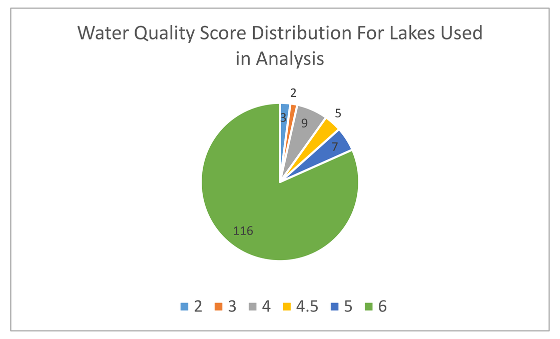
Sample Summary

The sample includes 169 lake watersheds from throughout Vermont, including several known "problem" lakes such as Carmi and Ticklenaked. Lakes whose watershed was easily extracted were given priority for inclusion to prevent overlap and clipping errors. The lakes/watersheds selected also had all or most of the variables of interest available. Some lacked a score, or some of the measured variables. Each was visited and sampled in at least one year between 2011 and 2014. As can be seen at right, the set is fairly evenly spread across the state.





The distribution of mean spring phosphorus measurents in the lakes.



The distribution of scores given to lakes by the Vermont ANR. Scores of 5 or 6 were qualified as having "good" water quality, and are not of great concern.

Results

Abandoning the Score: The scorecard water quality scores did not exhibit significant variation (the vast majority fell into the highest quality category, with only a few ranked "fair" or "reduced") and as such were not of interest for statistical analysis. The lakes are simply doing too well!
The main response variables of interest then became the raw nutrient measurements:

Table 1: Natural Log Transformed Average Spring Phosphorus Significant Linear Regressions

Explanatory Variable	R^2 (Highest to Lowest)
% Agricultural	0.286
LN Mean Elevation	0.178
LN % Agricultural	0.162
Mean Elevation	0.151
LN % Wetland	0.1
Mean Slope	0.088
LN Mean Slope	0.085
% Wetland	0.08
LN % Developed	0.035
% Developed	0.023
LN of Lake to Basin Size Ratio	0.023

Table 2: Average Spring Nitrogen Significant Linear Regressions

Explanatory Variable	R^2 (Highest to Lowest)
% Agricultural	0.187
LN % Agricultural	0.113
% Developed	0.081
LN Mean Elevation	0.056
LN % Developed	0.054
LN Mean Slope	0.052
Mean Elevation	0.041
Mean Slope	0.031
LN % Wetland	0.028

Table 3: LN Average Spring Chloride Significant Linear Regressions

Explanatory Variable	R^2 (Highest to Lowest)
LN % Developed	0.311
% Developed	0.283
% Agricultural	0.164
Road Density	0.149
LN Road Density	0.122
LN Mean Elevation	0.079
Mean Elevation	0.065
LN % Agricultural	0.055

Table 4: Average Spring pH Significant Linear Regressions

Explanatory Variable	R^2 (Highest to Lowest)
% Agricultural	0.208
LN % Agricultural	0.166
LN % Developed	0.09
Mean Elevation	0.067

Table 7: Clarification of units for variables incorporated in this study

Variables	Unit of original measure
Mean Elevation	US Feet
Mean Slope	Degree
Average Spring Phosphorus	Micrograms/Liter
Average Spring Nitrogen	Milligrams/Liter
Average Spring Chloride	Micrograms/Liter

Table 5: Results of multiple regressions using the enter method for each of the four spring response variables. The enter method was used, and all of the variables that exhibited individually significant linear relationships with a given variable were used.

Response Variable	Multiple Regression Equation (Enter Method)	Adjusted R^2
LN Average Spring Phosphorus	=(.032 * % Agricultural) + (.008* LN % Developed) +(.082 * LN % Wetland) + (.085 * LN Mean Slope) + (-.227 * LN Mean Elevation) + (.091 * LN Lake to Basin Ratio) + 4.023	.365
Average Spring Nitrogen	= (.008 * LN % Wetland) + (.013 * % Agricultural) + (-.002 * % Developed) + (-.01 * LN Mean Elevation) + (.005 * LN Mean Slope) + .300	.187
LN Average Spring Chloride	= (.027 * % Agricultural) + (.060 * % Developed) + (-.105 * LN Mean Elevation) + (107.265 * Road Density) + 1.496	.381
Spring pH	= (.004 * % Agricultural) + (6.684E-6 * Mean Elevation) + (.010 * LN % Developed) + 1.935	.194

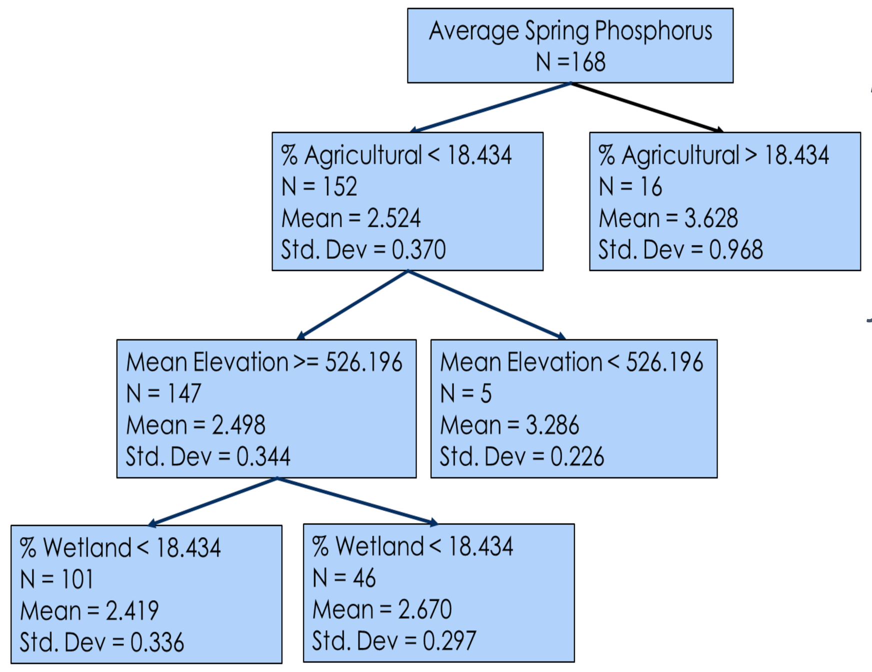


Figure 1: A partition tree showing the breakdown of sampled average spring phosphorus. Mean and Std. Deviation values represent phosphorus concentrations in micrograms per liter. Elevation value in feet.

Table 6: Results of multiple regressions run using the variables in the spring phosphorus partition, as well as interaction variables among them. The low/high distinction was determined using the partition cutoffs for each variable, and translated into binary fields to use in the regression.

Variables in Multiple Regression	Adjusted R^2 (Enter Method)
% Agricultural, Mean Elevation	.116
Mean Elevation, % Wetland	.027
% Agricultural, % Wetland	.113
% Agricultural, Mean Elevation, % Wetland	.110
% Agricultural, Mean Elevation, % Wetland, (LowAg*HighElev)	.158
% Agricultural, Mean Elevation, % Wetland, (HighElev*LowWetland)	.109
% Agricultural, Mean Elevation, % Wetland, (LowAg*HighElev), (HighElev*LowWetland)	.155
% Agricultural, Mean Elevation, % Wetland, (LowAg*HighElev), (HighElev*LowWetland), (LowAg*HighElev*LowWetland	.153

Acknowledgements

I would like to thank Leslie Matthews and Kellie Merrell of the Vermont Agency of Natural Resources for the raw lake water quality data and providing metadata for the variables, and Dan Homeier for providing the Vermont Agency of Natural Resources' spatial layers for the lakes and lake watersheds.

Funding provided by NSF EPS Grant #1101317