Impacts of Land Use on Water Quality in Minnesota

https://github.com/lhr12/HDA_Project

Keith Bollt, Jake Greif, Felipe Raby-Amadori, & Lindsay Roth

Abstract

Abstract tbd

<Information in these brackets are used for annotating the RMarkdown file. They will not appear in the final version of the PDF document>

Contents

List of Tables

List of Figures

\<Note: set up autoreferencing for figures and tables in your document\>

1 Research Question and Rationale

\<Paragraph detailing the rationale for your analysis. What is the significant application and/or interest in this topic? Connect to environmental topic(s)/challenge(s).\>

- Land use has a large impact on nutrient runoff into streams, lakes, and other water bodies
- Minnesota has wide variety of land uses. Inlcudes large urban centers, natural lands, and agricultural area.
- Nutrient management has been a challenge for states in the effort to control harmful algal blooms and coastal dead zones.
- Understanding the causes of nutrient problems will better inform management strategies.

\<Paragraph detailing your research question(s) and goals. What do you want to find out? Include a sentence (or a few) on the dataset you are using to answer this question - just enough to give your reader an idea of where you are going with the analysis.\>

Research questions:

1. What are the predictors of nutrients based on land use in watersheds in the state of Minnesota?

2. How do you characterize seasonal variation between the predictors of nutrients?

Goals: * Determine how land use, watershed size, and ecoregion explain variation in nutrient loading indicators. * Discern whether there a seasonal trends in nutrient loading indicators based on land use types, watershed size, and ecoregion. * Provide insight to inform decisions about nutrient managment practices based on land use types, watershed size, and ecoregion.

2 Dataset Information

<Information on how the dataset for this analysis were collected, the data contained in the dataset, and any important pieces of information that are relevant to your analyses. This section should contain much of same information as the README file for the dataset but formatted in a way that is more narrative.>

The data used in this analysis include data from the Lake Multi-Scaled Geospatial and Temporal Database (LAGOSNE) and the EPA ecoregion spatial datasets.

LAGOSNE is a collection of several data modules that contain information on lakes in the northern United States. The modules contain data from thousands of lakes in 17 states in the northeastern and midwestern United States, from Missouri to Maine. The dataset includes a complete list of all lakes bigger than 4 hectacres in the 17 state area, and water quality data on a large number of lakes, spanning every state.

Ecoregions are used by planning managers to understand the type of land use that occurs in different regions of the United States. There are different levels of ecoregions. Level 1 divides North America into 15 ecological regions, while Level IV offers fine ecological resolution for each state. This data was published by the U.S. EPA Office of Research and Development (ORD) - National Health and Environmental Effects Research Laboratory (NHEERL). For the purposes of our project, we selected Level III ecoregions because they appear to offer a descriptive narrative of the land use patterns of Minnesota without making a 'distinction without a difference'.

<Add a table that summarizes your data structure. This table can be made in markdown text or inserted as a `kable` function in an R chunk. If the latter, do not include the code used to generate your table.>

3 Exploratory Data Analysis and Wrangling

<Include R chunks for 5+ lines of summary code (display code and output), 3+ exploratory graphs (display graphs only), and any wrangling you do to your dataset(s).>

<Include text sections to accompany these R chunks to explain the reasoning behind your workflow, and the rationale for your approach.>

4   Analysis

<Include R chunks for 3+ statistical tests (display code and output) and 3+ final visualization graphs (display graphs only).>

<Include text sections to accompany these R chunks to explain the reasoning behind your workflow, rationale for your approach, and the justification of meeting or failing to meet assumptions of tests.>

- First we will create correlation plots in order to eliminate variables with a correlation greater than 0.8.
- Then we will run Shapiro-Wilkes tests to determine normality and the need for possible data transformations.
- After determining the distributions of the data, then we will generate mixed effect linear models with chlorophyll a and secchi depth as response variables, land use and watershed size as fixed effects, and ecoregion as a random effect.

Final figures will include: * 6 maps of the state, each showing the relationship between land use and both response variables. Ecoregion will be included as a base layer for each map. * Scatter plots showing the strongest relationships between land use and the response variables. * Table showing results of linear model.

5 Summary and Conclusions

<Summarize your major findings from your analyses. What conclusions do you draw from your findings? Make sure to apply this to a broader application for the research question you have answered.>

- Conclusions will include a discussion of our results within the context of MN nutrient management plan.