Proposal and Analysis Plan

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

October 28, 2019

What questions will your team address?

1. What are the predictors of nutrients in multiple watersheds within the state of Ohio?
2. Is there seasonal variation between the predictors of nutrients?

What hypotheses will your team address?

- Hypotheses should relate directly to your questions. Each numbered hypothesis should match up with the corresponding numbered question.
- There may be multiple working hypotheses for a single question. If this is the case, note each hypothesis as 1a, 1b. etc.

1a. Higher proportions of natural (forested and grassland) land use will be positively correlated water quality (indicated by Nitrogen and Phosphorus levels) due to higher rates of nutrient uptake in plants and lower outputs of nutrients in runoff.

1b. Higher proportions urban and agricultural land use will be negatively correlated with water quality due to higher rates of nutrient runoff.

1c. Higher amounts of atmospheric deposition will be negatively correlated with water quality due to increased precipitation and dryfall into watersheds.

2. Agricultural land use will be a more significant contributor to nutrient loading during fertilization seasons while urban land use will be the significant contributor to nutrient loading during the opposite seasons.

What dataset(s) will your team analyze?

LAGOS dataset

How will you set up and manage your project repository?

We will create a github project. Each team member will commit often with detailed commit messages.

Create a table of variables you will analyze.

- Column 1: Variable -total nitrogen, total phosphorus, iws land use 11-95, atmospheric nitrogen deposition
- Column 2: Units (if known)
-
- Column 3: Dependent (response) or independent (predictor) variable?
- tn and tp will be response, the others will be predictors

- Column 4: To which hypothesis(es) does this variable pertain? -all of our hypotheses will use these variables, with hypothesis 2 comparing seasons.

Variable	Units	Dependent_or_Independent	What_Hypotheses
tp	µg/L	Dependent	1+2
tn	µg/L	Dependent	1+2
Open Water	Land Use Classification	Independent	1+2
Urban	Land Use Classification	Independent	1+2
Bare rock/sand/clay	Land Use Classification	Independent	1+2
Forest	Land Use Classification	Independent	1+2
Scrub/shrub.	Land Use Classification	Independent	1+2
Orchards/vineyards/	Land Use Classification	Independent	1+2
Grasslands/herbaceous	Land Use Classification	Independent	1+2
Pasture/hay	Land Use Classification	Independent	1+2
Row crops	Land Use Classification	Independent	1+2
Wetlands	Land Use Classification	Independent	1+2
Season	Categorical	Independent	2

Note: You may not know all of the individual variables you plan to analyze at this point. It is sufficient to describe what type of variable you anticipate using (e.g., land cover) and decide on specifics later

What tasks will your team conduct?

Data acquisition

Determine the variables we need to grab from the LAGOS dataset and make dataframes including those variables.

Data exploration

We will make correlation plots, box/violin plots, qqnorm/qqline to test for normality.

Data wrangling

We'll create skinny datasets from the greater LAGOS database and filter the rows and columns we need for our analysis.

Data analysis and visualization

We will run linear models to determine which variables are the most significant predictors of nutrients. We will visualize our results in maps, scatterplots, tables.