

Regression Group Project Proposal

Group 4

10/20/2020

STAT 615-001

-Group members ;

Elizabeth Burch, Eli Schaub, Jiayi Sun, Kingsley Ofoegbu

Introduction

For our Group Project I, we will use Elizabeth's thesis research dataset of seeps and macroinvertebrate inhabitants of seeps to do a logistic regression.

A seepage spring is a type of shallow subterranean habitat. Like a spring, it is groundwater fed, but the flow is low, so it looks like a persistent wet spot in the woods. Seeps are typically found under forest canopy, covered in blackened leaves. The seep itself is rich in organic matter, but it is underlain by clay or another impermeable layer. The water temperature of the seep is more moderated (warmer in winter, cooler in summer) than that of surface streams.

Living in this habitat are aquatic amphipods and isopods with subterranean morphology: reduced or absent eyes and pigment, and elongated appendages. They are thought to come up to the seep exit to get organic carbon and nutrients. Many inhabit a very restricted range of just one or a handful of sites, and they may migrate between sites. These species have not been studied for very long, and little is known about their behavior, habitat requirements, or role in the forest ecosystem.

The study area for this dataset is National Parks Service parklands located between the Anacostia River and the D.C. border with Maryland. The parks have been surveyed and seep points georeferenced by AU students doing field work with Dr. David Culver. The amphipods sampled here include different species of *Stygobromus* and *Crangonyx*, and there are also isopods of the genus *Caecidotea*. The parklands are small and fragmented by roads. Seeps are very vulnerable to degradation; soil compaction, clear cutting, or the construction of impervious surfaces dissecting the habitat could threaten seep populations by causing both habitat loss and habitat fragmentation.

The research in this dataset will focus on which habitat quality and habitat isolation variables could be predictive of seeps being occupied by one or more shallow subterranean species. We will use logistic regression because the response variables are binary. Each seep is listed as inhabited or not, both overall and for each species. Data concerning habitat quality and degradation has been generated using ArcMap 10.7. The explanatory variables include data at each seep point of things like elevation, topographic wetness index, soil characteristics, and distance to nearest road, sewer outfall, or other inhabited seep. Because there are 11 soil characteristics variables available (such as water capacity, percent clay, permeability), a principal components analysis has been performed in order to choose just 2 or 3 factors to include in the analysis. We would like to focus on PCA in our Group Project II, so that we can introduce that topic and then use it as part of the logistic regression.

The work on this project will be split between group members who will each take the lead on different parts, such as exploratory analysis, report writing, and regression, but we will all help each other, especially with the data analysis. Group members will also create their own research questions and hypotheses about the dataset, within the overarching frame of exploring the effects of habitat quality, degradation, or isolation on seep occupancy.