**Background**

Populations of anadromous pacific salmon within California are declining and many populations are expected to disappear entirely within the next 100 years (Katz et al., 2013). For this project I focus on two of these species, coho salmon and steelhead. Of the two evolutionary significant units of coho in California, one is federally listed as threatened and the other is endangered (Moyle et al., 2017). For steelhead there are seven Distinct Population segments within the state, 5 of which are federally listed as threatened and one of which is endangered (Moyle et al., 2017). Numerous factors have contributed to their decline and provide ongoing threats, including, drought and other climate factors, hatchery interactions, and over-fishing (Brown et al., 1994). In some areas, habitat loss is thought to be the predominant mechanism of historical decline, driven by dams, agriculture, and logging (Brown et al., 1994). Urban land use has been linked to the decline of salmon overtime and has been shown to impact juvenile survival in the Pacific Northwest (Bilby & Mollot, 2008; Paulsen & Fisher, 2001). In California many salmon populations spawn within areas that are now considered protected, including lands which have previously been ecologically damaged. For example, Redwood Creek watershed, which now partially resides within Redwood National Park was previously extensively logged (Brown et al., 1994). I am interested in understanding how protected areas impact salmon populations in California, both in terms of population numbers and rate of change over time. Understanding this relationship can help managers improve land use practices to protect these species.

**Data**

To answer my questions, I used data from two different sources. For salmon population data I pulled data from the [California Monitoring Plan for salmon and steelhead](https://wildlife.ca.gov/Conservation/Fishes/Salmonid-Monitoring/CMP) (CMP), which is a joint effort between the California Department of Fish and Wildlife (CDFW) and NOAA to monitor and assemble data on anadromous salmon populations across California. For this project, I used estimates for spawning adult coho and steelhead, as this was available for many of the watersheds across a number of years. In addition to the count data, I used geospatial data from the CMP which indicated the watershed or sub-watershed extent that was monitored. To determine the amount of protected area within a watershed I used data from the [California Protected Areas Database](https://www.calands.org/)(CPAD) consisting of polygons of areas protected for open space use throughout California.

**Methods**

Data Prep

The initial stages of this analysis involved data prep and synthesis. To determine percent protected I combined the spatial polygon data for CMP watersheds with polygons from the CPAD. This allowed me to identify protected areas which overlapped with monitored watersheds, and calculate percent protected for each.

**Map

Description automatically generated**

**Figure 1:** Map of protected areas within watersheds

Analysis

**Analysis**

I used multiple OLS regression with an interaction effect. My dependent variable was adult population counts, which were collected annually across a number of watersheds. My independent variables were year and percent of watershed protected. I included an interaction term to allow the slope of the relationship between year and population count to vary based on percent protected. I ran two separate regressions for steelhead trout and coho salmon.

Population = B0 + B1year + B2 year\* %protected.

My null hypothesis in both scenarios is B1 = B2 = B3 = 0. My alternate hypothesis is B1≠B2 ≠B3 ≠ 0.

**Results**

**Discussion**

**OLS Violations**

My salmon population data does not appear to be linear in parameters for either species of interest. Salmon populations vary widely from year to year due to factors including ocean and climate conditions as well as hatchery releases within some watersheds, making it difficult to identify population trends (source). In addition, many monitored populations had 0 or 1 fish during my analysis time period. These sites likely represent areas which formerly had viable salmon populations, and now only have a few fish during years when there is overflow from nearby populations. All these factors may contribute to the lack of linearity in coho and steelhead populations over time, and likely cause the first assumption of OLS to be violated, causing issues with my analysis.

In addition to issues with the first assumption of OLS, my error does not appear to be normally distributed, violating the fourth assumption of OLS. For both species it is skewed to the right and has a long right tail. This indicates that OLS may not be the correct choice to analyze this data.

Due to these violations I cannot reject the null hypothesis.

**Study Limitations**

In terms of protected area data, my data had an extensive (roughly 30%) number of NAs in the year established column, making it unclear whether they existed during the whole study period. CPAD also uses a loose definition of protected and includes any lands that were maintained for open space purposes. It could be better to address this question using only protected areas which are specifically managed for biodiversity purposes. This type of data is available through [USGS](https://www.usgs.gov/programs/gap-analysis-project/science/protected-areas), however I opted not to use it due to even more extensive NAs in the year established column.

There are also data quality issues related to inconsistent monitoring of watersheds over the time of interest. Some watersheds were monitored over the whole time, while others were monitored intermittently or only over a few years. This could not be controlled for using cross sectional fixed effects due to the multicollinearity between watersheds at any time point and percent protected.

To further address my question, a better method would be to look at the relationship between protected areas and the rate of overwinter survival for juvenile salmonids. This would reduce issues related to non-linear population changes over time and normalize for populations with high numbers. It would also focus on the life-stage that is likely most impacted by freshwater habitat, and thereby inland protected areas. Unfortunately, overwinter survival data does not appear to be collected consistently across watersheds in California, which would make further data collection necessary for this analysis.