Predicting Salmon Viability in the Strait of Georgia

Sean Haight, Aycin Iplikci-Arodirik, Sayantan Roy, Andrew Staal

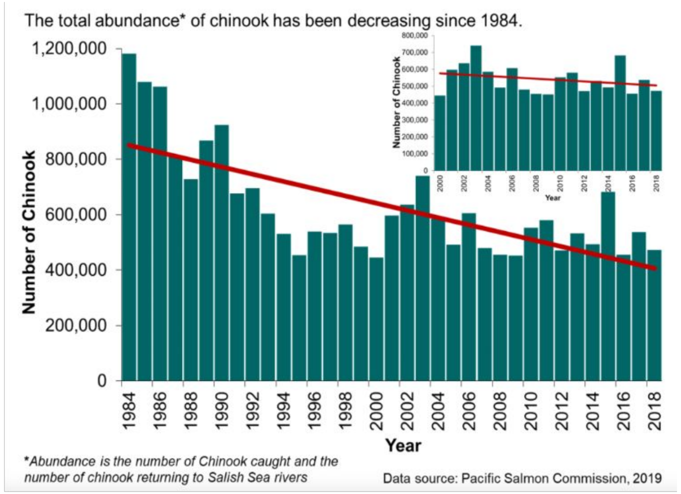
Young Chinook Salmon. Source:https://www.inaturalist.org/observations/79970429

The Strait of Georgia. Source:https://georgiastrait.org/issues/about-the-strait-2/

Motivation and Problem statement

Declining Salmon population on the Pacific coast since 1984

- Environmental impact: Disruption of energy and nutrient transfer between pacific ocean and freshwater rivers. Disruption of food chain for exotic salmon predators.
- Economic impact: Commercial and recreational salmon fisheries are worth ~$641 million annually, being a substantial employer in British Columbia.

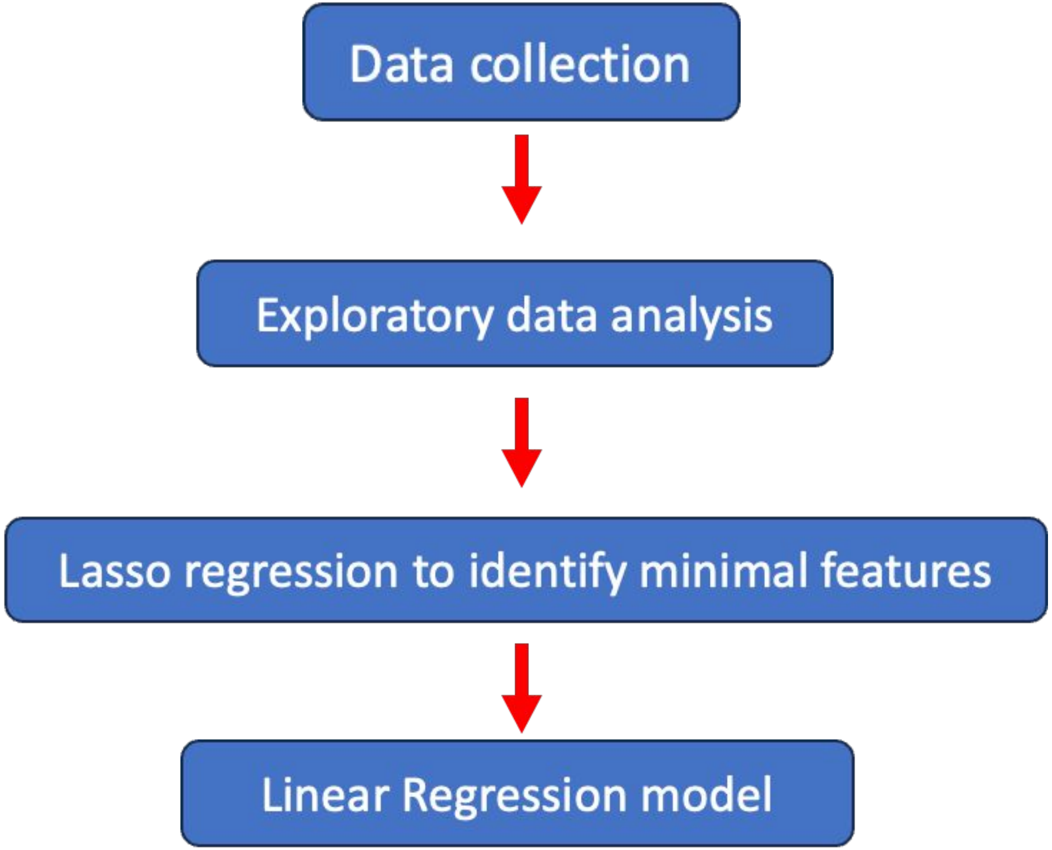


Problem: Identify environmental, predatory and human factors that have the largest impact on Salmon populations in the Strait of Georgia.

Stakeholders: Salmon fisheries in the Strait of Georgia, local indigenous peoples, environmental groups.

Goal: Identify a minimal set of features that have the most effects on the salmon viability and build regression model using estimated features. Will lead to development of strategies by stakeholders for better management of Salmon population.

Roadmap


Data collection
Exploratory data analysis
Lasso regression to identify minimal features
Linear Regression model

Data collection

Collection of data on relevant features for salmon population in SOG:

Relevant input features

Abiotic

a) Sea surface temperature
b) Sea level
c) Salinity
d) Fraser River flow rate
e) Climate indices (NPGO, PDO)
…

Biotic

a) Predation by Killer whales
b) Predation by Harbor Seals
c) Zooplankton availability

Anthropogenic

a) Fishery catches
b) Population in surrounding areas
c) Sea traffic
d) Pollution level and water quality
…

Output variable

Viability

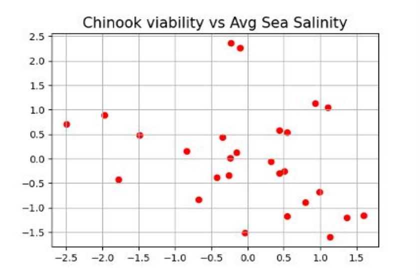
a) Cowichan Chinook
b) Harrison Chinook
c) Puntledge Chinook
d) Big Qualicum Coho

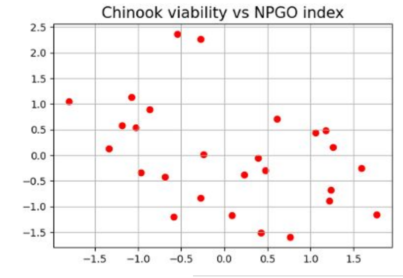
Challenges:

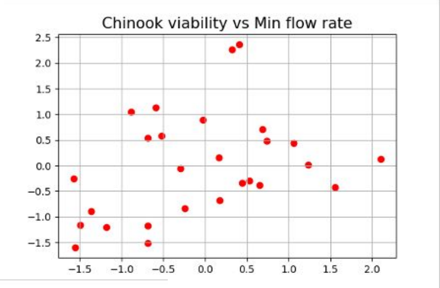
- No unique source for input features.
- Features may not overlap spatially and temporarily.
- Large number of features, that might be intercorrelated.

Exploratory Data Analysis

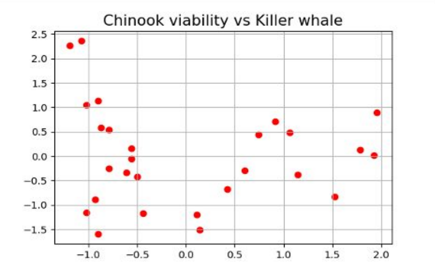
Abiotic features

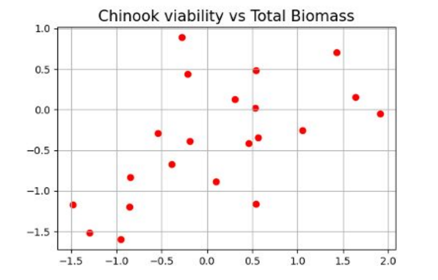




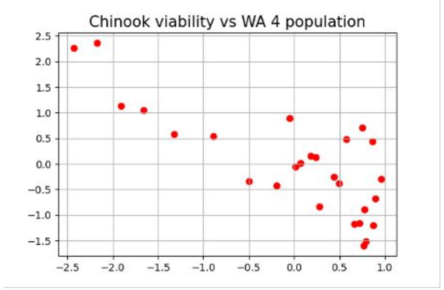


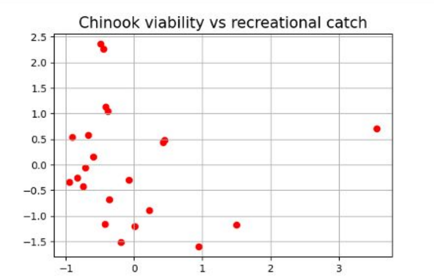
Biotic features

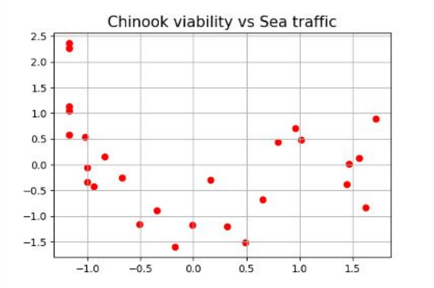




Anthropogenic features



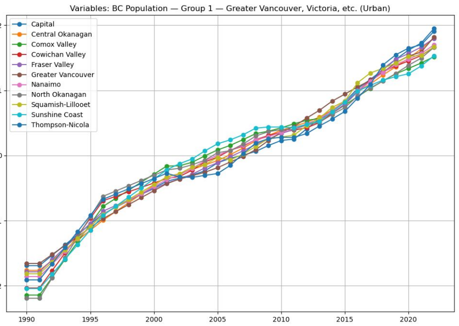


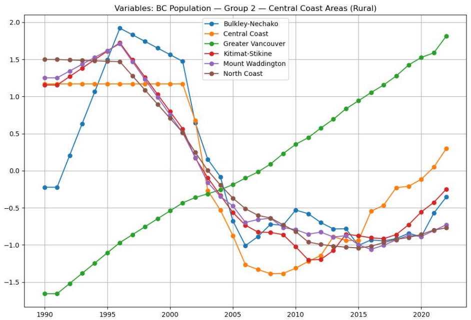


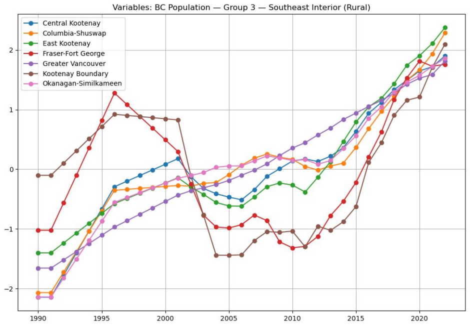
EDA+Lasso regression

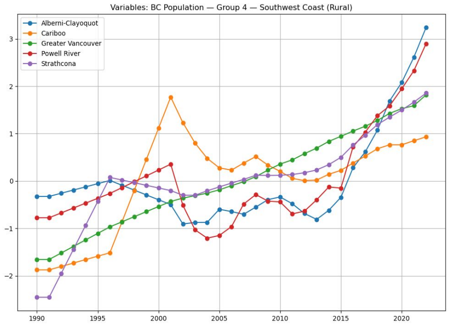
Primary stage: EDA revealed trends in small variable groups; preliminary significance ranking used lasso.

E.g. BC Regional District population trend groups:



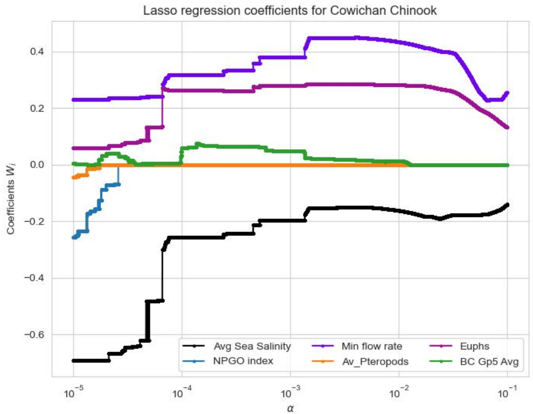






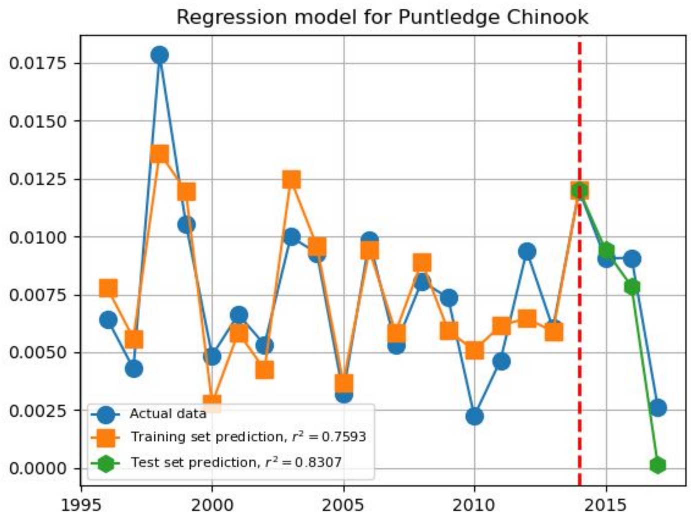
Secondary stage:

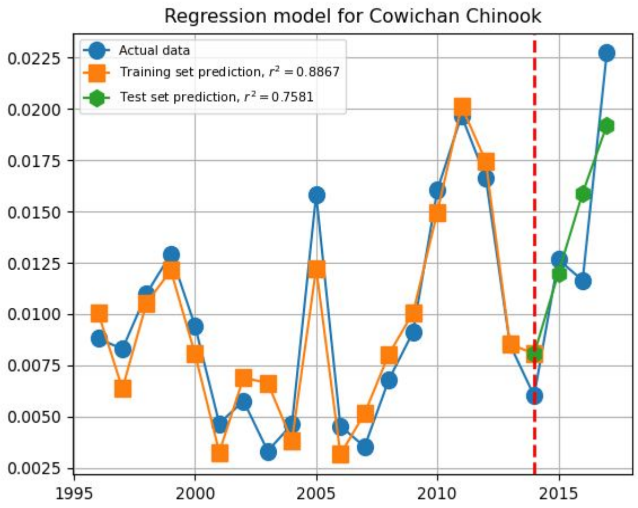
- All significant features combined into large dataframe with ~85 columns.
- For all 4 salmon types, 20 most significant features are identified with lasso.
- Features are identified by looping through lasso hyperparameter space and determining the top 25% features that survive.



Linear regression model

- The last three years of data (2015-2017) were separated as validation data.
- For each set of our 20 features, we formed subsets of size at most 5 features. We fit a linear model and chose the subset with the smallest MSE on the training data.





Future directions?

- Updated salmon survival rate and other data beyond 2017
    - Our survival data is from: Perry RI, Young K, Galbraith M, Chandler P, Velez-Espino A, Baillie S (2021) Zooplankton variability in the Strait of Georgia, Canada, and relationships with the marine survivals of Chinook and Coho salmon. PLoS ONE 16(1): e0245941.
    -
- Include Fraser River water quality measurements (no data pre-2000).
- Source more detailed water traffic data (AIS, WA ports).
- Inclusion of data on common pathogens and diseases in Salmon species.
- Other modeling approaches: e.g. polynomial regression with interaction terms, autoregression DFA.

Thanks to our wonderful mentors- Matt Osborne, Alec Clott, and Michael Darcy