

# Salmon Viability in the Strait of Georgia

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

Our goal is to determine which environmental factors impact salmon viability in the Strait of Georgia the most. To this end, we started by collecting data on salmon survival rates (our target variable) and on a wide variety of environmental & human factors including:

- Zooplankton Populations and their respective anomalies (major food source)
- Sea Surface Temperatures
- Flow rate from the Fraser River
- Harbor Seal and Killer Whale Populations (major predators)
- Reported fishery catch rates
- Sea Surface Salinities
- Sea level
- Fraser River water quality measurements
- The Pacific Decadal Oscillation (PDO), an index to measure long-term ocean fluctuation of the Pacific Ocean
- The North Pacific Gyre Oscillation (NPGO), an index to measure changes in ocean currents and circulation patterns
- Human population (BC Regional Districts/WA Counties)
- Tonnage of cargo (Port of Vancouver, Northwest Seaport Alliance)

The data collection is complicated by the fact that there is no unique repository for all the features considered here; this leads to both spatial and temporal non-overlaps between various features. Exploratory data analysis is done to eliminate spatial and temporal mismatch as much as possible, to reduce the number of highly correlated variables, and to select subsets of features indicative of processes occurring in the same geographical location and time window as our target variable (salmon population viability).

After processing the above data we plotted each of the above variables against our annual salmon viability data in order to see any clear relationships. We identified the following as possibly having a strong impact on salmon populations:

- Zooplankton Populations
- Sea Surface Temperatures
- Water salinity
- Recreational catch rates
- Pacific decadal oscillations
- Sea Level

Using the above data, we fit a Linear Regression model after using Lasso Regression in order to determine a list of 20 or fewer features that correlate most with salmon viability. Using these 20 features we used best subset selection to select five or fewer of these features that best predict salmon viability, reserving the last 3 years of our data solely for testing. We found the following most significant features for each Salmon species:

<b>Cowichan Chinook</b>	<b>Harrison Chinook</b>	<b>Puntledge Chinook</b>	<b>Big Qualicum Coho</b>
Average Sea Salinity	Sea Level at Campbell River	Cowichan Chinook Viability	Harbour Seal Populations
Medusae Average Biomass	Sea Surface Temperature in February	Sea Level at Cherry Point	Total Zooplankton Biomass
Euphs Average Biomass	Non-CalCops Average biomass	CalCops Average Biomass	Natantia Average Biomass
Fish Average Biomass	Larvaceae Average Biomass	Euphs Average Biomass	PolycatPelagic Average Biomass
Natantia Average Biomass	Siphonophorae Average Biomass	Repantia Average Biomass	WA Population (trend of coastal counties Island & Grays Harbor)

Note that the predominant features are different species of zooplankton biomass. In particular euphausiids play an important role in the modeling of both Cowichan Chinook and Puntledge Chinook. Thank you to our mentors Alec, Michael and Matt.