**METHODS**

**Data**

Data supporting this research are chiefly sourced from ADFG and NPS. ADFG conducts annual surveys of approximately 700 salmon streams in southeast Alaska in order to monitor pink salmon escapement for the purposes of managing the region’s commercial fishery. Survey data used in this study covers the period from 1960 to 2023. These streams, referred to by ADFG as index streams, are chosen because they are believed to reflect the natural condition of pink salmon populations in the region, meaning they have not been heavily impacted by human development. Index streams are subset into three sub-regions: (south southeast, north southeast (inside), and north southeast (outside). This last subregion, the neighborhood of Sitka and the Indian River, is the smallest, composed of 35 streams along the ocean-facing coasts of Baranof and Chichagof Islands as well as a few smaller islands in the vicinity. Pink salmon surveys are conducted via fixed-wing aircraft throughout the late summer spawning season. A random subset of those streams surveyed are selected each year for validation by foot counts to ensure the accuracy of ADFG’s aerial survey efforts (A. Dupuis, ADFG, personal communication, August 19, 2024). OBSERVERS RECORDED.

Due to the proximity of SJH and the accompanying potential of hatchery fish influencing abundances of wild fish, the Indian River is not one of ADFG’s index streams. Nonetheless, ADFG has still intermittently monitored pink salmon escapement in the Indian River in service of having the most complete data possible on pink salmon in the region. Data are available from Indian River from 1962-67, 1969, 1971-73, 1977-88, 1990, and from 1993-2024, meaning a total of 45 annual observations.

DESCRIPTIONS OF SPAWNER DATA (even and odd incl. IR)

This work also leverages data on rates of fish straying to Indian River from SJH, as well as rates of wild-origin fish returning to SJH as opposed to their stream of origin. Presumably, most wild fish straying to SJH would have originated in the Indian River, although it is not possible to determine conclusively through otoliths markings where wild-origin fish originated. These data are generated through three distinct processes: sampling of cost-recovery fish caught by purse-seiners, sampling of fish recovered at SJH as broodstock, and through sampling of deceased fish at Indian River by NPS biologists. Stray rate data is sparse in all three cases, with only 41 observations of stray rates (both in the park and at the hatchery) taken intermittently from 2011 onward and at inconsistent times throughout the season.

FIGURES ON STRAY RATES – probably need to get cleaned up

PDO data

**Methods**

The first stage of this research will attempt to assess whether abundances of pink salmon observed at Indian River are consistent with those abundances seen elsewhere in the region, with no attention paid to the existence of the hatchery. This is accomplished by comparing estimates of the underlying population of Indian River pink salmon to populations estimated in ADFG’s 35 north southeast (outside) index streams. Multivariate autoregressive state-space (MARSS) models are well suited for parsing long running time series survey data, and designed to mathematically distinguish observation error from error arising from stochasticity in the underlying state of a system. This feature is useful when parsing data collected by different observers operating in differing conditions (weather, time of day, time of year, etc.), where those conditions likely bias the observations being made. Furthermore, MARSS models allow for gaps in the observed time series, like those present in the Indian River data, while retaining the ability to make inferences about the underlying state (Holmes et al. 2012).

MARSS models are composed of a process (or state) model and an observation model, generically formulated as follows:

Observation equation: **yt** = **Zxt** + **a** + **Ddt**, + **vt**, where **vt**∼MVN(0,**R**)

Process/state equation: **xt** = **Bxt−1** + **u** + **Cct** + **wt**, where **wt**∼MVN(0,**Q**)

In the above, observed data in a given time period (**yt**) are used to estimate the underlying state of a system of interest (**xt**). In the observation equation, **Z** determines which time series observations in the observation equation will inform which underlying states in the process model. In the case of Indian River and the 35 index streams in its vicinity, specifying **Z** as a 36x36 identity matrix would allow each time series to be representative of a separate (in this case stream-level) state, while specifying **Z** as a 36x1 vector of 1s would force all time series in the data to represent samples from one comprehensive underlying state. **a** allows for trends in the observation process to be asserted or captured (depending on whether a is specified or estimated), and any correlation in observation error between time series is defined by **R**, the variance-covariance matrix. In the process equation, **B** defines the interaction between state estimates through time (i.e., **xt** and **xt−1**). If **B** is specified as an identity matrix, **u** then represents trends in the underlying state(s), otherwise **B** and **u** together determine the underlying mean and how quickly the time series would return to the mean following some perturbation. Finally, the correlation of process error is defined by **Q**. Finally, **ct** and **dt** are covariate data related to the state or sampling procedure respectively, with **C** and **D** capturing the effect of said covariates.

My research leverages MARSS models to estimate the annual abundance of pink salmon at Indian River and to then compare those estimates to pink salmon abundance in neighboring streams. These models include covariate controls for ocean climate and the observer of record for each time series observation (in north southeast (outside) index streams the observer is recorded in 77% of observations, with the rest being assigned an unknown observer). The specification of the **Z** is of particular importance to this work, as it defines the state against which Indian River pink salmon abundances will be compared. Thus far, the most parsimonious model as defined by AIC includes a 36x2 **Z** matrix which selects the Indian River as representative of one state and all other streams in the sub-region as representative of a second state. This is consistent with non-quantitative intuition if indeed the presence of SJH creates conditions that differ from a “natural” state.

The underlying state of Indian River pink salmon abundance seems to be higher than that of the surrounding region (Figure 1.3). This could be indicative of the impact of SJH pink salmon releases, but is far from conclusive in this regard. The next stage of this work will seek to isolate the effect hatchery operations have had on populations of pink salmon at Indian River.

The timeline of SJH potentially impacting pink salmon abundance at Indian River is reasonably straightforward: hatchery operations begin in 1975, meaning that given the short life cycle of pink salmon, the first returning adults would appear shortly thereafter (it is unclear from available documentation whether a brood of fish were released in 1975 or if that was the first year broodstock was taken from sources at Indian River and Starrigavan Creek). For the first three decades of its existence, SJH was permitted to propagate 1 million pink salmon eggs per year, with this number was increased to 3 million in 2010. This then defines three distinct time periods of SJH production: pre-1975 with no hatchery operations, 1975-2009 with low level operations, and post-2010 with high level operations (Stopha 2015).

This research seeks to take advantage of these distinct time periods by utilizing a difference-in-difference (DD) framework. This quasi-experimental approach is designed to assess the impact of an intervention (in this case the operations of SJH) on an outcome of interest (pink salmon abundance at Indian River). DD modeling compares a treatment group to a control group across two time periods, pre- and post- intervention. The general structure of a difference-in-difference model is as follows:

yi,t = β0 + β1Dtreatment + β2Dpost  + β3Dpost\*Dtreatment + ei,t

In the above, i indexes whether an observation (yi,t) relates to the treatment or control group, while t indexes whether that same observation comes from before or after the treatment is applied. Dtreatment is equal to 1 when yi,t is part of the treatment group and equal to 0 otherwise. Similarly Dpost is equal to 1 when yi,t is observed after treatment is applied, and equal to 0 when yi,t is observed before. These indicator variables allow for the estimation of a net effect of treatment on the treatment group (β3) by first controlling for differing characteristics in the treatment and control group (β1) and differences in the pre- and post-treatment time periods (β2). Additional control covariates may be included in order to further isolate the effects of interest. Crucially, input data must cover both a control and treatment group, as well as time periods before and after the intervention of interest. It is important to note that in this case, and indeed in almost every application of DD modeling, true controls in sense of those resulting from a randomized experimental design are unavailable. Instead, researchers must make due with treating what might be referred to as reference states as control groups. The assumption that these reference states are suitable proxies for randomized control groups is bolstered through the application of event studies which evaluate whether trends in control and treatment groups are mathematically similar in the pre-treatment period (Cunningham 2021).

Applying a DD framework to this research, observations of pink salmon abundances at Indian River are assigned to the treatment group while observations at neighboring ADFG index streams form the control group. Such a framework will also be used to evaluate pre- and post-treatment variance in year-to-year abundances. Pre- and post-1980 will be evaluated to indicate the treatment period (allowing a couple of years for hatchery operations to ramp up). A statistically significant estimate of β3 greater than 0 would indicate that SJH operations are exerting upward pressure on Indian River pink salmon abundance. A separate specification of the DD model will include an indicator variable related to periods pre- and post-2010, in an effort to assess the impact of heightened pink salmon production.

The last phase of this research will move beyond the effect of SJH hatchery operations and focus on the effect of stray rates themselves. It is important to consider straying both to and from Indian River, as both would influence the abundances observed in stream. As previously mentioned, observational data on rates of straying both from SJH to Indian River and of wild-origin fish at SJH are sparse. They do however provide enough information to inform a suite of hypothetical straying scenarios.

Rates of straying to Indian River by hatchery fish observed across different sampling efforts vary from 0% to 80%. Generally, hatchery strays are observed in greater proportions at the river early in the spawning season (August and early September). Hatchery-origin strays at Indian River chiefly show otolith markings from SJH, although a small handful of pink salmon from the Port Armstrong Hatchery at the southern tip of Baranof Island have also been observed. These PAH origin fish make up less than 5% of all hatchery-origin fish observed at Indian River. Rates of wild-origin fish being picked up by SJH’s cost recovery efforts or as broodstock likewise vary considerably, from as low as 5% up to 100%. Wild-origin fish are more likely to stray to SJH later in the season, although this temporal pattern is not as pronounced at the hatchery as it is at Indian River.

Based on the ranges of straying observed, weighted tranches of straying will be defined for both hatchery-to-wild and wild-to-hatchery straying. Low, medium, and high stray rate tendencies will be defined for each of these groups, each with a mean and variance from which synthetic annual data can be drawn. When interacted, this will create 9 scenarios to be evaluated, each related to a distinct combination of levels of straying from both sources (i.e., IR-high/SJH-high, IR-low/SJH-high, IR-medium/SJH-low, etc.). These 9 scenarios will act as the basis for MARSS models focusing on the Indian River directly. The estimates of interest in these specifications will not be of the underlying state, but instead those elements of the **C** which relate to the synthetic rates of straying. Because this modeling will be focused only on Indian River, additional covariates such as stream flow and temperature may be included which are unavailable at other ADFG pink salmon index streams. These models are intended to outline the impacts of different levels of straying on abundances of pink salmon at Indian River, with the hope that future survey efforts might generate more robust stray rate data (consistent in timing at both sites) with which these results can be utilized.

**Results**

The numbers of pink salmon returning to Indian River are higher in most years than what might be expected for a given stream in the north southeast (outside) sub-region (Figure 1.3). This does not necessarily mean that SJH is driving this behavior, and indeed, if all streams were modeled as representative of individual underlying states, there will likely be a range of behaviors distributed around some general trend. Estimating β3 in a DD framework will provide more conclusive evidence of the impact of SJH operations. Additionally, evaluating the impact of the 2010 increase in SJH pink salmon production may provide insight as to how the impact of hatchery operations on wild populations scales. Finally, evaluation of various combinations of straying scenarios will create a ‘road map’ of what conditions are likely to unnaturally influence abundances of pink salmon at Indian River.