

# Challenges and Tools in the Assessment and Management of Pacific Salmon Fisheries

by

Benjamin A. Staton

A Dissertation submitted to the Graduate Faculty of  
Auburn University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Auburn, Alabama  
May 5, 2019

Keywords: Fisheries management, Bayesian inference, decision analysis

Copyright 2019 by Benjamin A. Staton

Approved by

Matthew J. Catalano, AFFILIATION  
Asheber Abebe, AFFILIATION  
Lewis G. Coggins, Jr., AFFILIATION  
Conor P. McGowan, AFFILIATION

## Abstract

I'm going to write an abstract to go here. This is the first paragraph of the dissertation abstract, which will talk about chapter 1..

This is the second paragraph of the dissertation abstract, which will talk broadly about chapter 2.

This is the third paragraph of the dissertation abstract, which will talk broadly about chapter 3.

This is the fourth paragraph of the dissertation abstract, which will talk broadly about chapter 4.

## Acknowledgments

Here is where I will thank everyone.

Matt, Lew, Brendan, Mike, Sam, AL-HPC folks, Steve, Nick, Zach, Janessa. Family and Michelle. Folks at the lab. RStudio staff.

## Table of Contents

Abstract . . . . .	ii
Acknowledgments . . . . .	iii
List of Figures . . . . .	v
List of Tables . . . . .	vi
Preface . . . . .	1
1    Assessment Approaches for Single-Species, Mixed-stock Pacific Salmon Fisheries:	
Empirical and Simulation-Evaluation Applications . . . . .	2
Abstract . . . . .	2
1.1    Introduction . . . . .	3
1.2    Methods . . . . .	6
1.2.1    Analytical approach . . . . .	6
1.2.2    Study system . . . . .	7
1.2.3    Data sources . . . . .	7
1.2.4    Regression-based models . . . . .	12
1.2.5    The full state-space model . . . . .	15
1.2.6    Alternate state-space models . . . . .	20
1.2.7    Parameter estimation . . . . .	21
1.2.8    Simulation-estimation analysis . . . . .	21
1.2.9    Metrics of model performance . . . . .	22
1.3    Results . . . . .	22
1.4    Discussion . . . . .	23
Bibliography . . . . .	31

## List of Figures

1.1	Visualization of how different types of heterogeneity in substock productivity and size influence the shape of trade-offs in mixed-stock salmon fisheries. Solid black lines are the case where stock types are split evenly among large/small and productive/unproductive stocks. Dotted black lines are the case where all small stocks are productive and all large stocks are unproductive, and dashed lines are the opposite (i.e., all big stocks are productive). (a) Equilibrium aggregate harvest and proportion of substocks overfished plotted against the exploitation rate (b) value of the biodiversity objective (0 = all stocks overfished) plotted against the value of harvest (the long term proportion of the aggregate MSY attained). Notice that when all big stocks are productive (dashed lines), the trade-off is steeper, i.e., more harvest must be sacrificed in order to ensure a greater fraction of substocks are not overfished. . . . .	27
1.2	The frequency of escapement sampling for each substock sampled in the Kuskokwim River. Black points indicate years that were sampled for substocks monitored with a weir and grey points indicate years sampled for substocks monitored with aerial surveys. The vertical black line shows a break where > 50% of the years were monitored for a stock. . . . .	28
1.3	The relationship between spatially-expanded aerial survey estimates and weir counts during the same years and substocks as described by (1.9). Notice the uncertainty expressed in the predictor variable; this was included in the analysis by incorporating both the spatial (Section 1.2.3.1.1) and temporal (Section 1.2.3.1.2) expansions in a single model fitted using Bayesian methods. . . . .	29
1.4	Estimated Chinook salmon escapement for substocks within the Kuskokwim River drainage. "Drainage-wide" refers to the aggregate population estimates provided by a maximum likelihood run reconstruction model. "This analysis" refers to the estimated portion of the aggregate run included in this analysis (not all tributaries are monitored). . . . .	30

## List of Tables

1.1	Description of the various indices used in the description of the state-space model. $n_t$ is the number of years observed for the most data-rich stock. . . . .	24
1.2	The estimated spatial expansion factors for the various aerial survey projects described in Section 1.2.3.1.1. In cases where multiple projects were flown in a larger subdrainage, the expanded counts were summed to obtain an estimate for the larger subdrainage, as indicated by the footnotes. . . . .	25
1.3	The estimated temporal expansion parameters for converting spatially-expanded aerial counts to estimates of subdrainage-wide escapement abundance each year.	26

## Preface

I used bookdown (Xie, 2015) to make this document.

## Chapter 1

# Assessment Approaches for Single-Species, Mixed-stock Pacific Salmon Fisheries: Empirical and Simulation-Evaluation Applications

## **Abstract**



## 1.1 Introduction

Many salmon populations in large drainage systems are harvested primarily in a relatively small spatial area and are managed as a single stock (i.e., the concept of a “mixed-stock fishery”). However, these “stocks” are instead stock-complexes, in which the aggregate stock is comprised of several (and sometimes, many) substocks. These substocks are known to show differences in genotypic (Templin et al., 2014), phenotypic (e.g., morphology; Hendry and Quinn, 1997), behavioral (e.g., run timing; Clark et al., 2015; Smith and Liller, 2017), and life history (i.e., age-at-maturation, Blair et al., 1993) characteristics that are the result of adaptations to local environments after many generations of high spawning site fidelity and reproductive isolation from conspecifics in other tributaries. It has been widely proposed that maintaining this diversity of local adaptation (hereafter, “biodiversity”) is favorable both from ecosystem and exploitation perspectives (i.e., the statistical dampening of random variability in a system made up of many additive random processes, otherwise known as the “portfolio effect”; Schindler et al., 2010, 2015).

This level of variability in substock-specific characteristics can ultimately lead to heterogeneity in productivity among the substock components (Walters and Martell, 2004). Productivity is the ability of a population to replace itself after harvesting, often represented for salmon populations as the maximum number of recruits (future migrating adults before harvest) per one spawner, which (due to density-dependent survival) is attained at very low numbers of spawners (hereafter,  $\alpha$ ). Substocks  $j$  with higher  $\alpha_j$  values can sustain greater exploitation rates ( $U$ ) than those with smaller  $\alpha_j$  values, in fact,  $\alpha_j$  can be expressed in terms of the exploitation rate that maximizes sustained yield from substock  $j$  (Schnute and Kronlund, 2002):

$$\alpha_j = \frac{e^{U_{\text{MSY},j}}}{1 - U_{\text{MSY},j}}. \quad (1.1)$$

Given that there is likely some level of heterogeneity in  $\alpha_j$  and  $U_{\text{MSY},j}$  among individual substocks, the logical conclusion is that in a mixed-stock fishery where  $U_t$  is common among all substocks, some weaker substocks must be exploited at  $U_t > U_{\text{MSY},j}$  in order to fish the more productive substocks at  $U_{\text{MSY},j}$ . This of course implies a trade-off, and in some cases it might be necessary to over-exploit some substocks in order to maximize harvest (Figure 1.1, Walters and Martell 2004).

Before these trade-offs are considered by managers in a well-informed way, the shape and magnitude of the trade-off must first be quantified as shown in Figure 1.1. Figures like this are generated using the estimated productivity and carrying capacity of all (or a representative sample) of the substocks within a mixed-stock fishery. These quantities are obtained using a spawner-recruit analysis, which involves tracking the number of recruits that were produced in each brood year (i.e., parent year) by the number of fish that spawned in the same calendar year and fitting a curve to the resulting pattern. The spawner-recruit literature is extensive, but primarily focuses primarily on assessing single populations as opposed to substock components (but see the work on Skeena River sockeye substocks Walters et al., 2008; Korman and English, 2013). In my mind, this is due to two factors:

- (1) the data to conduct well-informed substock-specific spawner-recruit analyses are often unavailable (20 – 30 years of continuous spawner and harvest counts/estimates and age composition for each substock) and
- (2) management actions in large mixed-stock fisheries may not be precise enough to deliberately direct more harvest activity towards particular substocks, so deriving substock-specific estimates could be of little utility.

This proposed chapter will pertain to salmon systems for which there is a reasonable amount of data available for a significant portion of the substocks and in situations where spawner-recruitment analysis estimates are desired for each.

The methods to fit spawner-recruit models can be grouped into two broad categories: time-independent error models (i.e., Clark et al., 2009) and state-space (i.e., time series) models (Fleischman et al., 2013; Su and Peterman, 2012). The independent error models typically take on a regression analytical method, and is thus subject to substantial pitfalls when dealing with data with inherent time-dependent properties and oftentimes large amounts of observation error (Walters and Martell, 2004). The state-space class of models captures the process of recruitment events leading to future spawners while simultaneously accounting for variability in the biological and measurement processes that gave rise to the observed data (de Valpine and Hastings, 2002; Fleischman et al., 2013). Including this level of additional model complexity comes at computational costs, as these models are well-suited for Bayesian inference with Markov Chain Monte Carlo (MCMC) methods (Newman et al., 2014, Ch. 4), but has been shown to reduce bias in estimates in some circumstances [Su and Peterman (2012); walters-martell-2004].

In the Kuskokwim River Chinook salmon fishery, there has been recent interest in considering biodiversity to explicitly inform the drainage-wide escapement goal. In conducting the spawner-recruit analysis to inform such policy analyses, it will be difficult to determine which method is appropriate, given many possible model structures, sparse data, and unknown sampling biases. Before strong inferences can be made from the ultimate trade-off analyses of interest, the performance of the estimation models used to parameterize them needs to be evaluated, as well as the appropriate level of model complexity needed to address the problem with sufficient accuracy. In this final chapter, I evaluate the performance of a range of assessment models for mixed-stock salmon fishery via simulation-estimation. The objectives will be to:

- (1) develop a set of varyingly-complex multi-stock versions of the state-space spawner-recruit models that have been rapidly gaining popularity, particularly in Alaska (Walters and Martell, 2004; Su and Peterman, 2012; Fleischman et al., 2013; Staton et al., 2017),

- (2) determine the sensitivity of trade-off conclusions to assessment model complexity (including those obtained using linear regression approaches) using empirical data from Kuskokwim River Chinook salmon substocks, and
- (3) test the performance of the assessment models *via* simulation-estimation.

## 1.2 Methods

### 1.2.1 Analytical approach

This analysis will be conducted in both an empirical and a simulation-estimation framework to evaluate the sensitivity and performance of assessment strategies for the mixed-stock assessment problem in Pacific salmon fisheries. First, all assessment methods will be fitted to observed data from the Kuskokwim River substocks ( $n_j = 13$ ) for the empirical objective. Then, a hypothetical system will be generated with known dynamics and will be comprised of several age-structured substocks. Then, these hypothetical populations will be sampled per a realistic sampling scheme (i.e., frequency of sampling, appropriate levels of observation variance, etc.). Each of the assessment models will be fitted to the resulting data sets, and the management quantities  $U_{\text{MSY}}$  and  $S_{\text{MSY}}$  (both on an aggregate and substock basis) will be calculated from the resulting estimates. The estimated quantities will then be compared to the true driving parameters and will be summarized and model performance will be compared among a set of competing estimation models. Inference from the simulation regarding which assessment models perform the best can then be used to justify an appropriate level of model complexity for this problem when applied to the Kuskokwim and systems like it. I will begin by describing the data sources, followed by the estimation models assessed in this study, and then provide other necessary details regarding the simulation-estimation analyses.

### 1.2.2 Study system

The models developed in this chapter will be fitted to empirical data from substocks of the Kuskokwim River. Figure ?? shows the Kuskokwim River drainage, as well as the many assessment projects used to inform the 13 substocks used in this analysis. The fishery characteristics are discussed in Sections xxx, xxx, xxx, but it can very well be described as a mixed-stock fishery, both for multiple salmon species (Chinook, chum *O. keta*, and sockeye *O. nerka*) and for multiple substocks of the same species. Fish originating and returning to the various tributaries enter through the bulk of the fishery as a mixed-stock, though Chinook salmon stocks traveling to the headwaters have been illustrated to enter the main stem earliest in the summer migration (Smith and Liller, 2017) so a limited ability to direct harvest towards or away from these stocks is possible manipulating the front portion of the fishery (reference to chapter 3). It is acknowledged that the assessment program does not sample all tributaries within the Kuskokwim River.

### 1.2.3 Data sources

All data for this analysis are available to the public, and came primarily from the Arctic-Yukon-Kuskokwim Database Management System (AYKDBMS)<sup>1</sup> maintained by the Alaska Department of Fish and Game (ADF&G). Cases in which other data sources were necessary are highlighted in the description below, e.g., the telemetry data needed to perform the expansion of aerial survey counts described in Section 1.2.3.1 below.

#### 1.2.3.1 Substock escapement

Escapement count data for this analysis were informed predominately by the ADF&G Kuskokwim River salmon escapement monitoring program, the details of which have been most-recently documented in (Head and Liller, 2017). The data set available spans 20 different

---

<sup>1</sup><http://www.adfg.alaska.gov/CommFishR3/WebSite/AYKDBMSWebsite/Default.aspx>

escapement monitoring projects (6 weirs and 14 aerial surveys) and 42 calendar years from 1976 – 2017. For substocks monitored *via* weir, observed substock  $j$  escapement in year  $t$  ( $S_{obs,t,j}$ ) was taken to be the total estimated weir passage each year. Substocks monitored *via* aerial survey needed special care, however. Surveys have been flown only once per year on a relatively small fraction of each tributary system (Head and Liller, 2017), resulting in these data being indices of escapement rather than estimates of total escapement. This analysis required estimates of total escapement to each substock however, because this would allow calculation of biological reference points that are expressed in terms of the scale of the population (e.g., the spawner abundance that is expected to produce maximum recruitment;  $S_{MAX,j}$ ), rather than as a rate (i.e.,  $U_{MSY,j}$ ). Note that if only estimates of  $U_{MSY}$  were desired, no accounting for the partial count would be necessary.

The approach I developed to estimate total escapement from single-pass aerial surveys involved two main steps:

- (1) Mapping the distribution of detected telemetry-tagged Chinook salmon against distribution of the aerial survey counts. This comparison allowed for a spatial expansion to estimate how many salmon would have been counted if the entire tributary had been flown.
- (2) Obtaining and applying a temporal correction factor for the problem of counting a dynamic pool at one point in its trajectory. This correction factor was based on the relationship between paired weir and aerial counts on  $n = 3$  of the systems in the analysis.

### 1.2.3.1.1 Spatial expansion

The core of the the spatial expansion estimator was the assumption:

$$\frac{A_{f,t,i}}{T_{f,t,i}} = \frac{A_{u,t,i}}{T_{u,t,i}}, \quad (1.2)$$

where the quantities  $A$  and  $T$  represent fish and tags, respectively in flown ( $A_f$  and  $T_f$ ) and unflown ( $A_u$  and  $T_u$ ) reaches in year  $t$  and for aerial survey monitoring  $i$ . This assumption states that the ratio of spawners per one tagged spawner is the same between flown and unflown river sections at the time of the aerial index count and the aerial telemetry flights. Equation (1.2) and can be rearranged as:

$$A_{u,t,i} = A_{f,t,i} \frac{T_{u,t,i}}{T_{f,t,i}}. \quad (1.3)$$

If  $T_{u,t,i}$  is further assumed to be a binomial random variable with time-constant success parameter  $p_i$ , then:

$$T_{u,t,i} \sim \text{Binomial}(p_i, T_{u,t,i} + T_{f,t,i}). \quad (1.4)$$

Here,  $p_i$  represents the probability that a tagged fish in the spawning tributary monitored by project  $i$  was outside of the survey flight reach at the time of the aerial telemetry flight. When (1.4) is rearranged to put  $p_i$  on the odds scale, then:

$$\psi_i = \frac{p_i}{1 - p_i}. \quad (1.5)$$

Estimated expansion factors  $\psi_i$  and  $p_i$  are shown in Table 1.2. The odds value  $\psi_i$  can be substituted for the division term in (1.3) which gives:

$$A_{u,t,i} = A_{f,t,i} \psi_i. \quad (1.6)$$

To obtain the total number of fish that would have been counted had the entire subdrainage been flown ( $\hat{A}_{t,i}$ ), the components can be summed:

$$\hat{A}_{t,i} = A_{f,t,i} + A_{u,t,i}. \quad (1.7)$$

Substitution of (1.6) into (1.7) and factoring gives the estimator:

$$\hat{A}_{t,i} = A_{f,t,i}(1 + \psi_i). \quad (1.8)$$

The spatial expansion model was integrated with the temporal expansion model described below into a single model fitted in the Bayesian framework.

### 1.2.3.1.2 Temporal Expansion

The temporal expansion model was necessary to convert from the one-pass index scale to the substock escapement scale: it was a temporal correction. The temporal expansion I developed operated by first regressing  $n = 16$  observations of paired weir count ( $W_i$ ) and spatially-expanded aerial counts ( $\hat{A}_i$ ; given by (1.8)) on the same tributary systems ( $n = 3$ ) in the same years:

$$\begin{aligned} W_i &= \beta_0 + \beta_1 \hat{A}_i + \varepsilon_i, \\ \varepsilon_i &\stackrel{\text{iid}}{\sim} N(0, \sigma_W) \end{aligned} \quad (1.9)$$

The estimated coefficients  $\hat{\beta}_0$  and  $\hat{\beta}_1$  (Table 1.3) were then applied to tributary systems with an aerial count but not a weir count:

$$S_{obs,t,j} = \hat{\beta}_0 + \hat{\beta}_1 \hat{A}_{t,j} \quad (1.10)$$

For stocks that had both weirs and aerial surveys, the weir count was used as  $S_{obs,t,j}$  as opposed to using the expansion in (1.10). The fitted relationship is shown in Figure 1.3.



### 1.2.3.2 Aggregate harvest

Harvest estimates for the Kuskokwim River are available at the drainage-wide scale only, and were obtained each year by subtracting the drainage-wide estimates of total run and escapement (Liller et al., 2018). Because the escapement data used here do not encompass all the substocks within the Kuskokwim River system, it was necessary to remove some portion of the total harvest that was produced by stocks not included in this analysis. First, I calculated the observed exploitation rate of the drainage-wide Kuskokwim River Chinook salmon stock ( $U_{obs,t}$ ) by dividing the total harvest by the total run each year. I then made the assumption that monitored and unmonitored substocks have received the same exploitation rates, in which case total harvest accounted for in this analysis harvest was be obtained as:

$$H_{obs,t} = \frac{S_{obs,t}U_{obs,t}}{1 - U_{obs,t}}, \quad (1.11)$$

which can be derived from the definition of the exploitation rate ( $U = \frac{H}{S+H}$ ). This step was embedded within the same Bayesian model that encompassed the spatial and temporal aerial survey expansions such that uncertainty in these steps could be propagated through the entire analysis.

Note that  $S_{obs,t}$  and  $H_{obs,t}$  do not have  $j$  subscripts denoting particular substocks: this indicates that they are aggregate quantities summed across all substock components. In cases where substock-specific harvest was desired (i.e., in reconstructing the substock-specific brood tables for fitting individual regression relationships; Section 1.2.4.3),  $H_{obs,t,j}$  was obtained using (1.11) by substituting  $S_{obs,t,j}$  in for  $S_{obs,t}$ .

### 1.2.3.3 Age composition

Age composition data were necessary to reconstruct brood tables for age-structured salmon populations (see Section 1.2.4.3). Age data used in this analysis came from the

ADF&G standardized age, sex, and length sampling program operated at the weir projects. All sampled fish that were not aged successfully were discarded as were samples corresponding to the rare ages of 3 (0.2 and 1.1) and 8 (1.6 and 2.5) such that only fish successfully aged as between 4 and 7 were included. It is possible that older or younger fish may have the systematic tendency to return early or late in the run, and this could introduce biases if age sampling was not conducted proportionally to fish passage throughout the season. To adjust for this possibility, a weighted-average scheme was applied to obtain the age composition estimates for each substock and year with data. Daily age samples were stratified into two-week strata and strata-specific proportions-at-age were calculated. These strata-specific age compositions were then averaged across strata within a year and stock weighted by the number of Chinook salmon estimated to have passed the weir in each stratum. The total number of fish successfully aged for each year and substock was retained for data-weighting purposes for the state-space models (see Section 1.2.5.3).

#### 1.2.4 Regression-based models

Two regression-based approaches to estimating Ricker (1954) spawner-recruit parameters in the multi-stock case were assessed: (*a*) a single mixed-effect regression model with random intercepts and (*b*) independent regression models. A description of each method is provided in the sections that follow.

##### 1.2.4.1 Mixed-effect linear regression

The Ricker (1954) spawner-recruit model can be written as:

$$R_y = \alpha S_y e^{-\beta S_y + \varepsilon_y} \tag{1.12}$$

where  $R_y$  is the total recruitment produced by the escapement  $S_y$  in brood year  $y$ ,  $\alpha$  is the maximum recruits-per-spawner (RPS),  $\beta$  is the inverse of the escapement that produces maximum recruitment ( $S_{\text{MAX}}$ ), and  $\varepsilon_y$  are independent mean zero normal random variables attributed solely to environmental fluctuations. Primary interest lies in estimating the population dynamics parameters  $\alpha$  and  $\beta$  as they can be used to obtain biological reference points off of which sustainable policies can be developed. This function is increasing at small escapements and declining at large ones, though can be linearized:

$$\log(\text{RPS}_y) = \log(\alpha) - \beta S_y + \varepsilon_y, \quad (1.13)$$

allowing for estimation of the parameters  $\log(\alpha)$  and  $\beta$  in a linear regression framework using the least squares method (Clark et al., 2009). This relationship is nearly always declining, implying a compensatory effect on survival (i.e., RPS) with reductions in spawner abundance (Rose et al., 2001). Regression-based methods for estimating spawner-recruit parameters are well known to be fraught with two primary issues:

- (1) ignoring the intrinsic time linkage whereby brood year recruits (part of the response variable) make up the escapement for the one or more future brood years (the predictor variable), which then produce the future recruits (response variables) and
- (2) ignoring the fact that escapement and harvest are often measured with substantial error.

The first issue is known as the “time-series bias”, and is known to chronically cause positive biases in  $\alpha$  and negative biases in  $\beta$ , causing the same directional biases in  $U_{\text{MSY}}$  and  $S_{\text{MSY}}$ , respectively (i.e., spuriously providing too aggressive harvest policy recommendations; Walters, 1985). The second is known as the “errors-in-variables bias” and is known to cause an apparent scatter which inserts additional variability that commonly-used regression estimators do not

account for (Ludwig and Walters, 1981). Though these methods have been known for their problems for over 30 years, they are still somewhat widely used (Korman and English, 2013).

It is not difficult to conceive a multi-stock formulation of this model by including substock-specific random effects on the intercept  $[\log(\alpha)]$ :

$$\begin{aligned}\log(\text{RPS}_{y,j}) &= \log(\alpha_j) - \beta_j S_{y,j} + \varepsilon_y, \\ \log(\alpha_j) &= \log(\alpha) + \varepsilon_{\alpha,j}, \\ \varepsilon_{\alpha,j} &\sim \text{N}(0, \sigma_\alpha).\end{aligned}\tag{1.14}$$

It does not make sense to include substock-level random effects on the slope, given that  $\beta$  is a capacity parameter related to the compensatory effect of resource limitation experienced by juveniles, likely in the freshwater environment (i.e., amount of habitat as opposed to quality of habitat). Fitting the individual substock models in this hierarchical fashion allows for the sharing of information such that the more intensively-assessed substocks can help inform those that are more data-poor.

#### 1.2.4.2 Independent regression models

The mixed-effect model may have the benefit of sharing information to make some substocks more estimable, but it should also have the tendency to pull the extreme  $\alpha_j$  (those in the tails of the hyperdistribution) toward  $\alpha$ . This behavior may not be preferable for policy recommendations, as it should tend to dampen the extent of heterogeneity estimated in  $\alpha_j$ . For this reason, independent regression estimates for each substock will also be obtained (i.e., the full fixed effects model) for evaluation.

#### 1.2.4.3 Brood table reconstruction

An important consideration in the use of the regression-based method is in how  $RPS_{y,j}$  is obtained. Only the states  $S_{y,j}$  are ever directly observed;  $R_{y,j}$  is observed (for Chinook salmon) over for calendar years as not all fish mature and make the spawning migration at the same age. Thus, in order to completely observe one  $RPS_{y,j}$  outcome, escapement must be monitored in year  $y$  and escapement, harvest, and age composition must be monitored in the subsequent years  $y + 4$ ,  $y + 5$ ,  $y + 6$ , and  $y + 7$ . It is evident that missing one year of sampling (which is a common occurrence; Figure 1.2) can lead to issues with this approach. Only completely observed  $RPS_{y,j}$  data were used for this analysis, with the exception of missage count data. For substocks with no age composition data (i.e., those monitored *via* aerial survey), the average age composition each year across substocks that have data was used to reconstruct  $RPS_{y,j}$ , but will be provided only for years with escapement sampling for substock  $j$ . Only substocks with  $\geq 3$  completely-observed pairs of  $RPS_{y,j}$  and  $S_{y,j}$  were included for model fitting.

#### 1.2.5 The full state-space model

Four versions of the state-space formulation were assessed. As three versions were simplifications of the full model, the full model is presented completely here and the changes resulting in the other three model structures are described in the subsequent section.

The state-space formulation of a multi-stock spawner recruit analysis developed and evaluated here is an extension of various single-stock versions (e.g., Fleischman et al. 2013). Walters et al. (2008) used a similar model using maximum likelihood methods to provide estimates of  $>50$  substocks in the Skeena River drainage, British Columbia. The model presented here will be fitted in the Bayesian mode of inference using the program JAGS (Plummer, 2017), and will relax certain assumptions made by Walters et al. (2008) such as the important notion of perfectly-shared recruitment residuals (i.e., anomalies – deviations

from the expected population response). It will also have the ability to relax the assumption of constant maturity schedules across brood years. See Table 1.1 for a description of the index notation, in particular note the difference between the brood year index  $y$  and the calendar year index  $t$ .

The state-space model can be partitioned into two submodels: (a) the process submodel which generates the true states of  $R_{y,j}$  and the resulting calendar year states (e.g.,  $S_{t,j}$ ) and (b) the observation submodel which fits the observed data to the true states. The model is fitted to three primary data sources:

- (1)  $S_{obs,t,s}$  – escapement estimates from the  $n_j$  substocks with data observed over  $n_t$  calendar years (some of which may be missing observations; see Section 1.2.3.1),
- (2)  $H_{obs,t}$  – calendar year estimates of aggregate harvest summed across all substocks included in the analysis (see Section 1.2.3.2), and
- (3)  $q_{obs,t,a,j}ESS_t$  – the calendar year age composition (relative contribution of each age class to the total run) for all substocks that have this information (see Section 1.2.3.3).

The proportion at age is contained in the  $q$  variable, and the weighting factor (number of fish sampled for age data in year  $t$ ) is contained in the  $ESS_t$ .

Note that this method allows for missing calendar year observations and does not require excluding brood year recruitment events that are not fully observed as was done for the regression-based models (see Section 1.2.4.3).

#### 1.2.5.1 Process submodel: brood year processes

The recruitment process operated by producing a mean prediction from the deterministic Ricker (1954) relationship (1.12) for  $n_y$  brood years for each of the  $n_j$  substocks. From these deterministic predictions, autocorrelated process variability is added to generate the true realized state. To populate the first  $n_a$  calendar year true states with recruits of each age  $a$ , the first  $a_{max}$  brood year expected recruitment states were not linked to a spawner abundance

through (1.12) (because the  $S_y$  component was not observed), but instead were assumed to have a constant mean equal to unfished equilibrium recruitment (where non-zero  $S_j$  produces  $R_j = S_j$  when unexploited and in the absence of process variability):

$$\bar{R}_{y,j} = \frac{\log(\alpha_j)}{\beta_j}, \quad (1.15)$$

where  $\bar{R}_{y,j}$  is the expected (i.e., deterministic) recruitment in brood year  $y$  from substock  $j$  with Ricker parameters  $\alpha_j$  and  $\beta_j$ . The remaining  $n_y - a_{max}$  brood years had an explicit time linkage:

$$\bar{R}_{y,j} = \alpha_j S_{t,j} e^{-\beta_j S_{t,j}}, \quad (1.16)$$

where  $t = y - a_{max}$  is the  $t^{\text{th}}$  calendar year index in which the escapement produced the recruits in the  $y^{\text{th}}$  brood year index.

From these deterministic predictions of the biological recruitment process, autocorrelated lag-1 process errors were added to produce the true realized states:

$$\log(R_{y,1:n_j}) \sim \text{MVN}(\log(\bar{R}_{y,1:n_j}) + \omega_{y,1:n_j}, \Sigma_R), \quad (1.17)$$

where

$$\omega_{y,1:n_j} = \phi (\log(R_{y-1,j}) - \log(\bar{R}_{y-1,j})), \quad (1.18)$$

where  $R_{y,1:n_j}$  is a vector of true recruitment states across the  $n_j$  stocks in brood year  $y$ ,  $\omega_{y,1:n_j}$  is the portion of the total process error attributable to serial autocorrelation,  $\phi$  is the lag-1 auto-correlation coefficient, and  $\Sigma_R$  is a covariance matrix representing the white noise portion of the total recruitment process variance. The covariance matrix  $\Sigma_R$  was estimated such that each substock was assigned a unique variance and covariance with each other substock. The

multivariate normal errors were on the log scale so that the variability on  $R_{y,j}$  was lognormal, which is the most commonly used error distribution for describing spawner-recruit data sets (Walters and Martell, 2004). Further, the multivariate normal was used as opposed to  $n_j$  separate normal distributions so that the degree of synchrony in brood-year recruitment deviations (i.e., process errors) among substocks could be captured and freely estimated.

The maturity schedule is an important component of age-structured spawner-recruit models, as it determines which calendar years the brood year recruits  $R_{y,j}$  return to spawn (and be observed). Recent state-space spawner-recruit analyses have accounted for brood year variability in maturity schedules as Dirichlet random vectors drawn from a common hyperdistribution characterized by a mean maturation-at-age probability vector ( $\pi_{1:n_a}$ ) and an inverse dispersion parameter ( $D$ ) (see Fleischman et al., 2013; Staton et al., 2017) for implementation in JAGS), and the same approach was used here with maturity schedules shared perfectly among substocks within a brood year. Brood year-specific maturity schedules were treated as random variables such that:

$$p_{y,a} \stackrel{\text{iid}}{\sim} \text{Dir}(\pi_{1:n_a} D). \quad (1.19)$$

where  $p_{y,a}$  is the probability a fish spawned in brood year  $y$  matures at age  $a$ . While there is almost certainly some level of between-substock variability in average maturity schedules, I made many attempts to estimate it and include it in these models, but all efforts resulted in either (1) nonsensical maturity estimates, (2) systematic residual patterns among substocks with and without age composition data, or (3) required auxiliary (i.e., never observed) information for substocks that do not have age composition information in order to fit. This result indicates to me that the variability is not estimable from the available data when embedded into this model. I think it is reasonable to expect that brood year deviations should be similar between substocks given that the factors that set the probability of maturing at



age are likely linked to growth and mortality conditions in the ocean part of the life-cycle, in which case all substocks would experience similar conditions.

### 1.2.5.2 Process submodel: calendar-year processes

In order to link  $R_{y,j}$  with calendar year observations of escapement from each substock,  $R_{y,j}$  will be allocated to calendar year runs:

$$N_{t,j} = \sum_{a=1}^{n_a} R_{t+n_a-a,j} p_{t+n_a-a,a}, \quad (1.20)$$

where  $N_{t,j}$  is the run abundance in calendar year  $t$  from substock  $j$ . The harvest process will be modeled using a freely estimated annual exploitation rate ( $U_t$ ) time series for fully-vulnerable substocks:

$$H_{t,j} = N_{t,j} U_t, \quad (1.21)$$

and escapement will be obtained as:

$$S_{t,j} = N_{t,j} (1 - U_t). \quad (1.22)$$

For both simulation and empirical analyses, all substocks were assumed to have been fished at the same rate (i.e., all stocks have been equally vulnerable to harvest). Though this may not be a terribly realistic assumption, I have made attempts at reconstructing vulnerability based on what is known about substock specific run timing and the harvest timing of the lower river fishery and found that on average differences in vulnerability were negligible.

The quantities  $N_t$  and  $S_t$  aggregated among all substocks can be obtained by summing within a  $t$  index across the  $j$  indices. Calendar year age composition for each substock was obtained by dividing an age-structured vector of the aggregate run at year  $t$  and age  $a$  by the total aggregate run in year  $t$ .

### 1.2.5.3 Observation submodel

Three data sources were be used to fit the model: observed (estimated) escapement from each substock ( $S_{obs,t,j}$ ) with assumed known coefficients of variation (CV), total harvest arising from the aggregate stock ( $H_{obs,t}$ ) with assumed known CV, and the age composition of substocks with age composition (the substocks monitored using weirs;  $n = 6$  for the Kuskokwim River) each calendar year ( $q_{obs,t,a,j}$ ) (which has associated effective sample size  $ESS_{t,j}$  equal to the number of fish successfully aged for substock  $j$  in year  $t$ ). The CVs were converted to lognormal standard deviations:

$$\sigma_{\log} = \sqrt{\log(\text{CV}^2 + 1)}, \quad (1.23)$$

and used in lognormal likelihoods to fit the time series  $S_{t,j}$  to  $S_{obs,t,j}$  and  $H_t$  to  $H_{obs,t}$ . Calendar year age composition was fitted using parameter vectors  $q_{t,1:n_a,j}$  and observed vectors of  $(q_{obs,t,1:n_a,j} ESS_{t,j})$ .

### 1.2.6 Alternate state-space models

Three alternate formulations of the state-space model will be evaluated, and all are simplifications of the full model described above regarding the structure of (1) the covariance matrix on recruitment residuals and (2) the maturity process. The simplest model will not include brood year variability in maturity schedules and  $\Sigma_R$  will be constructed by estimating a single  $\sigma_R^2$  and  $\rho$ , and populating the diagonal elements with  $\sigma_R^2$  and off-diagonal elements with  $\rho\sigma_R^2$ . One drawback of constructing  $\Sigma_R$  this way is that  $\rho < -0.05$  for a  $13 \times 13$  covariance matrix results in positive-indefiniteness, which is prohibited by JAGS. Thus, a constraint is required to maintain  $-0.05 \leq \rho < 1$  to prevent the sampler from crashing. In one intermediate model, brood year maturation variability will be included but the covariance matrix will be constructed as in the simplest model. In the other intermediate model, brood year variability

in maturation will not be included but the covariance matrix will be fully estimated as in the full model.

Certainly there are other model structures that could have been evaluated. For example, additional of complexity to the AR(1) time series model on recruitment anomalies could have been added (or it could have been removed entirely), other methods could be devised for obtaining the expected recruitment in the first  $a_{max}$  brood years (though I tried many), or different variance structures for white noise recruitment anomalies. These are all examples of different model structures that an analyst may choose, but I needed to limit the boundaries of the model evaluation for computational and inferential reasons. I chose the simple *versus* complex comparisons for recruitment and maturity variability because I think these are the two key areas where an analyst will question if they have enough data to fit the model. In other words, these are two key structural uncertainties where it is important to know if the complex versions are estimable with a reasonable amount of data.

### **1.2.7 Parameter estimation**

All parameter estimation was conducted in the Bayesian framework using JAGS (Plummer, 2017) implemented using R (R Core Team, 2018) and the R package `{jagsUI}` (Kellner, 2017) for interfacing with JAGS from R, as well as the `{coda}` (?) package and a package of my own creation `{codaTools}` (Staton, 2018).

### **1.2.8 Simulation-estimation analysis**

#### **1.2.8.1 Operating model: Biological submodel**

Given that the state-space model is a much more natural model of this system (which has intrinsic time series properties) than the regression-based versions, it will be used as the foundation operating model (i.e., state-generating model). The biological submodel will be

more complex than the most complex estimation model – namely in regards to the maturity schedule, which will have a modest level of substock variability but with highly correlated brood year variability. In order to serve as the state-generating model for the simulation, the state-space model needs only to be populated with true parameters, initial states, and a harvest control rule. I will use a fixed escapement policy with implementation error to ensure the data time series are generated with patterns consistent with realistic exploitation patterns (the policy will not be updated as more data are available). I will generate  $n_j = 12$  substocks with different parameters  $U_{\text{MSY},j}$  and  $S_{\text{MSY},j}$  which (as a starting point) will be informed from random draws from the joint posterior distribution of 13 substocks from the Kuskokwim River drainage.

#### 1.2.8.2 Operating model: Observation submodel

For a given set of simulated true states, a set of observed states ( $S_{\text{obs},t,j}$ ,  $H_{\text{obs},t}$ ,  $q_{\text{obs},t,a}$ ) will be generated by adding sampling error to each year, which will represent the value that would be observed if the sampling project operated that year. Observation errors in escapement and harvest estimates will be lognormal and multinomial for the age composition, as assumed in the state-space estimation model. Frequency of sampling on each substock (i.e., simulated data collection) will be set to approximately mimic the Kuskokwim River historical monitoring program. Approximately half of the substocks will have age composition data sampled in the same years as escapement, and aggregate harvest ( $H_{\text{obs},t}$ ) will be available every year in each simulation.

#### 1.2.9 Metrics of model performance

### 1.3 Results

I found some stuff.

## 1.4 Discussion

Here's what it means.

Table 1.1: Description of the various indices used in the description of the state-space model.  $n_t$  is the number of years observed for the most data-rich stock.

Index	Meaning	Dimensions
$y$	Brood year index; year in which fish were spawned	$n_y = n_t + n_a - 1$
$t$	Calendar year index; year in which observations are made	$n_t$
$j$	Substock index	$n_j$
$a$	Age index; $a = 1$ is the first age; $a = n_a$ is the last age	$n_a$
$a_{min}$	The first age recruits can mature	1
$a_{max}$	The last age recruits can mature	1

Table 1.2: The estimated spatial expansion factors for the various aerial survey projects described in Section 1.2.3.1.1. In cases where multiple projects were flown in a larger subdrainage, the expanded counts were summed to obtain an estimate for the larger subdrainage, as indicated by the footnotes.

<b>Aerial Survey</b>	$p_i$	$1 + \psi_i$
<b>Kisaralik</b>	0.59 (0.42 – 0.75)	2.46 (1.72 – 4.04)
<b>Salmon (Aniak)<sup>1</sup></b>	0.04 (0.01 – 0.12)	1.04 (1.01 – 1.14)
<b>Aniak<sup>1</sup></b>	0.41 (0.37 – 0.47)	1.71 (1.58 – 1.87)
<b>Kipchuk<sup>1</sup></b>	0.09 (0.04 – 0.17)	1.1 (1.04 – 1.21)
<b>Holokuk</b>	0.37 (0.23 – 0.53)	1.59 (1.3 – 2.12)
<b>Oskawalik</b>	0.44 (0.29 – 0.6)	1.79 (1.4 – 2.52)
<b>Holitna</b>	0.79 (0.75 – 0.83)	4.78 (4.04 – 5.73)
<b>Cheeneetnuk<sup>2</sup></b>	0.25 (0.16 – 0.38)	1.34 (1.18 – 1.61)
<b>Gagaryah<sup>2</sup></b>	0.08 (0.02 – 0.19)	1.08 (1.02 – 1.24)
<b>Salmon (Pitka Fork)<sup>3</sup></b>	0.4 (0.3 – 0.5)	1.66 (1.42 – 2.01)
<b>Bear<sup>3</sup></b>	0.05 (0 – 0.22)	1.05 (1 – 1.28)
<b>Upper Pitka Fork<sup>3</sup></b>	0.62 (0.48 – 0.75)	2.62 (1.92 – 4)

<sup>1</sup> Tributaries of the Aniak subdrainage

<sup>2</sup> Tributaries of the Swift subdrainage

<sup>3</sup> Tributaries of the Pitka subdrainage

Table 1.3: The estimated temporal expansion parameters for converting spatially-expanded aerial counts to estimates of subdrainage-wide escapement abundance each year.

Parameter	Estimate
$\hat{\beta}_0$	1.9 (-60.71 – 62.4)
$\hat{\beta}_1$	2.3 (1.76 – 2.85)
$\hat{\sigma}_W$	4992.15 (3376.54 – 7565.08)



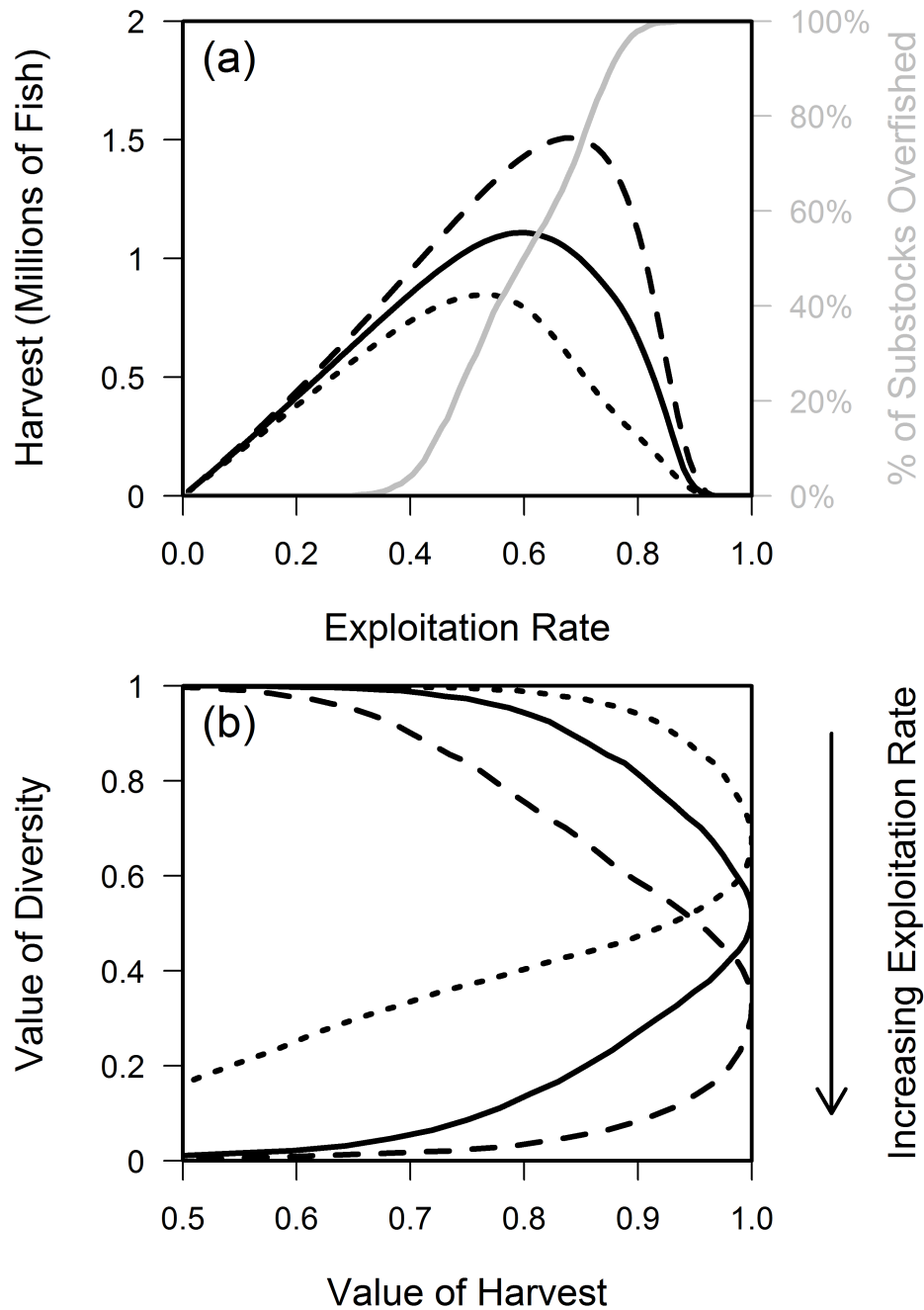


Figure 1.1: Visualization of how different types of heterogeneity in substock productivity and size influence the shape of trade-offs in mixed-stock salmon fisheries. Solid black lines are the case where stock types are split evenly among large/small and productive/unproductive stocks. Dotted black lines are the case where all small stocks are productive and all large stocks are unproductive, and dashed lines are the opposite (i.e., all big stocks are productive). (a) Equilibrium aggregate harvest and proportion of substocks overfished plotted against the exploitation rate (b) value of the biodiversity objective (0 = all stocks overfished) plotted against the value of harvest (the long term proportion of the aggregate MSY attained). Notice that when all big stocks are productive (dashed lines), the trade-off is steeper, i.e., more harvest must be sacrificed in order to ensure a greater fraction of substocks are not overfished.

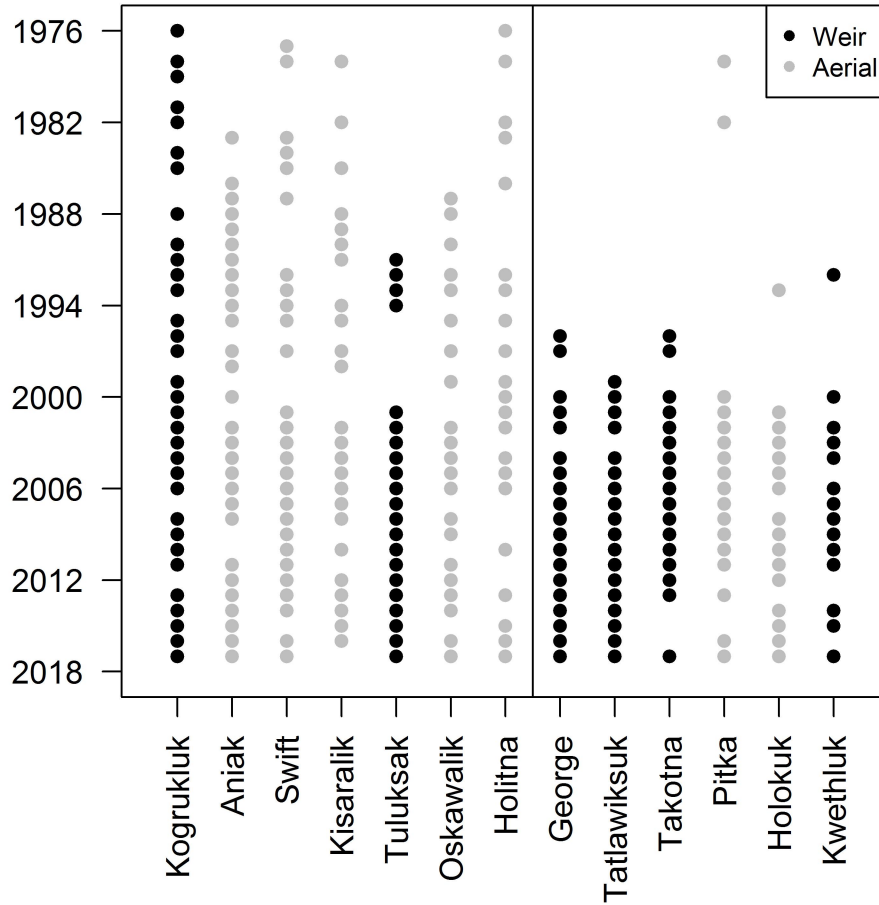


Figure 1.2: The frequency of escapement sampling for each substock sampled in the Kuskokwim River. Black points indicate years that were sampled for substocks monitored with a weir and grey points indicate years sampled for substocks monitored with aerial surveys. The vertical black line shows a break where  $> 50\%$  of the years were monitored for a stock.

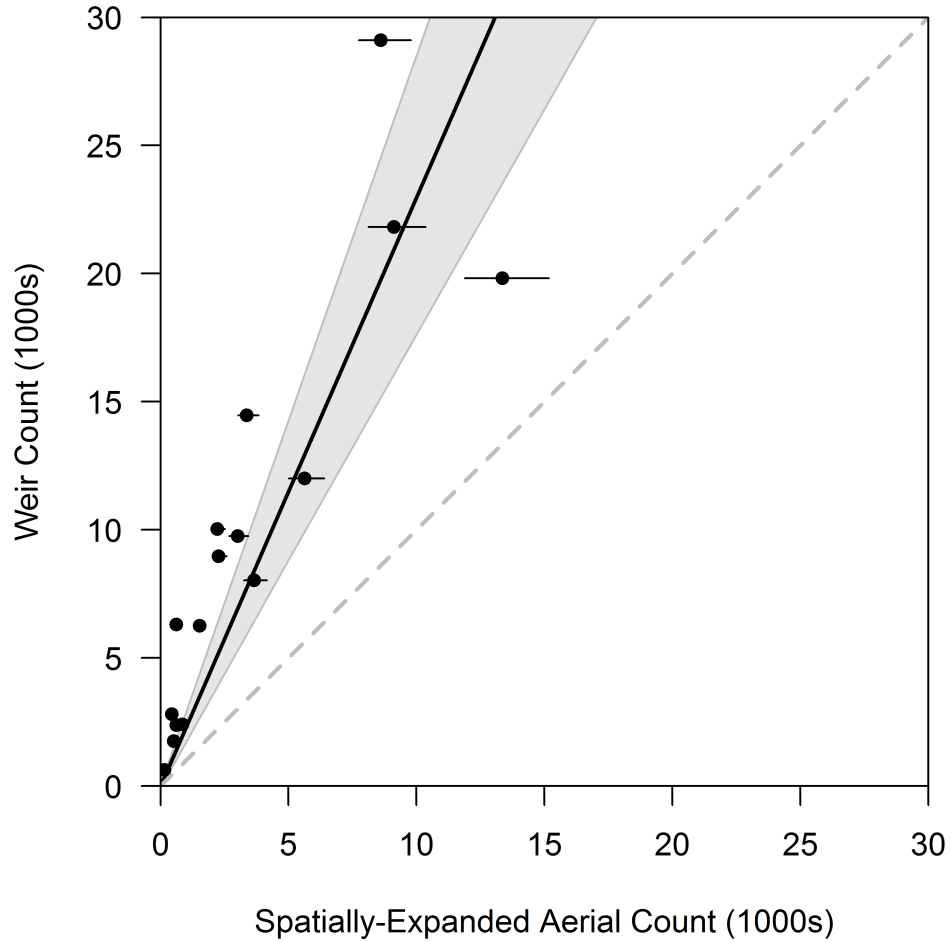


Figure 1.3: The relationship between spatially-expanded aerial survey estimates and weir counts during the same years and substocks as described by (1.9). Notice the uncertainty expressed in the predictor variable; this was included in the analysis by incorporating both the spatial (Section 1.2.3.1.1) and temporal (Section 1.2.3.1.2) expansions in a single model fitted using Bayesian methods.

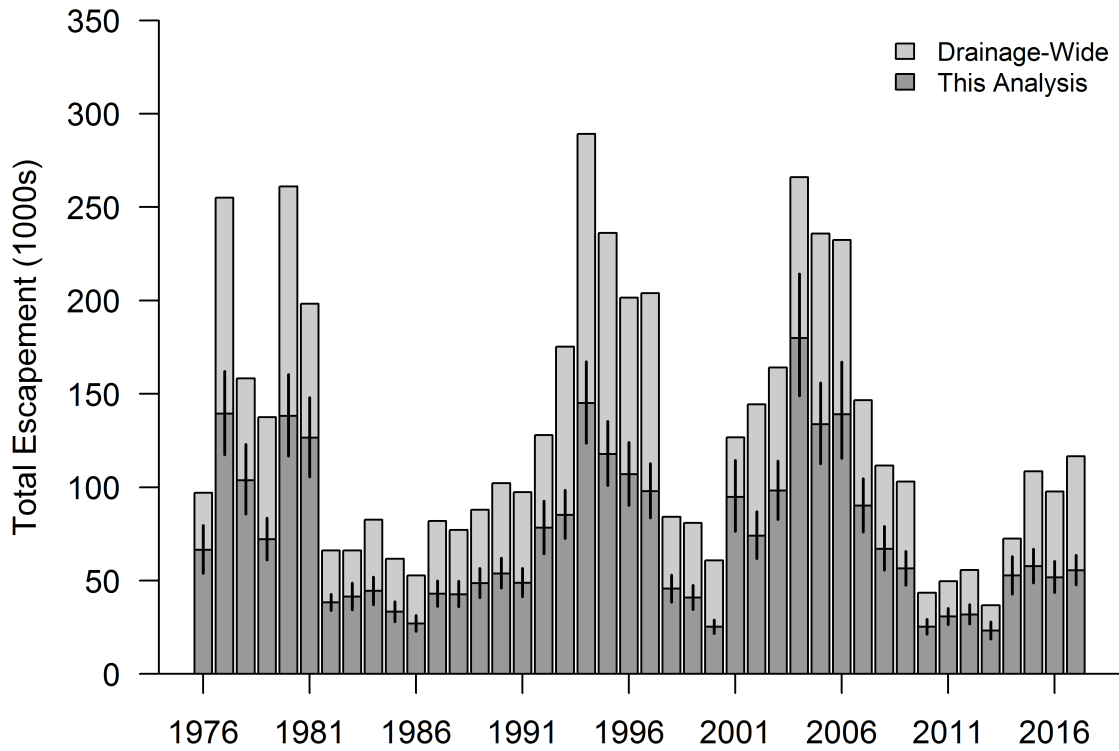


Figure 1.4: Estimated Chinook salmon escapement for substocks within the Kuskokwim River drainage. "Drainage-wide" refers to the aggregate population estimates provided by a maximum likelihood run reconstruction model. "This analysis" refers to the estimated portion of the aggregate run included in this analysis (not all tributaries are monitored).

## Bibliography

- Blair, G. R., Rogers, D. E., and Quinn, T. P. (1993). Variation in life history characteristics and morphology of sockeye salmon in the Kvichak River system, Bristol Bay, Alaska. *Transactions of the American Fisheries Society*, 122.
- Clark, R. A., Bernard, D. R., and Fleischman, S. J. (2009). Stock-recruitment analysis for escapement goal development: a case study of pacific salmon in alaska. In Krueger, C. C. and Zimmerman, C. E., editors, *Pacific Salmon: Ecology and Management of Western Alaska's Populations*, American Fisheries Society Symposium 70, pages 743–757, Bethesda, MD.
- Clark, S. C., Tanner, T. L., Sethi, S. A., Bentley, K. T., and Schindler, D. E. (2015). Migration timing of adult Chinook salmon into the Togiak River, Alaska, watershed: Is there evidence for stock structure? *Transactions of the American Fisheries Society*, 144(4):829–836.
- de Valpine, P. and Hastings, A. (2002). FITTING POPULATION MODELS INCORPORATING PROCESS NOISE AND OBSERVATION ERROR. *Ecological Monographs*, 72(1):57–76.
- Fleischman, S. J., Catalano, M. J., Clark, R. A., and Bernard, D. R. (2013). An age-structured state-space stock–recruit model for pacific salmon (*Oncorhynchus* spp.). *Canadian Journal of Fisheries and Aquatic Sciences*, 70(3):401–414.
- Head, J. and Liller, Z. W. (2017). Salmon escapement monitoring in the Kuskokwim Area, 2016. Fishery Data Series 17-29, Alaska Department of Fish and Game, Anchorage, AK.
- Hendry, A. P. and Quinn, T. P. (1997). Variation in adult life history and morphology among Lake Washington sockeye salmon (*Oncorhynchus nerka*) populations in relation to habitat features and ancestral affinities. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(1):75–84.
- Kellner, K. (2017). *jagsUI: A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses*. R package version 1.4.9.
- Korman, J. and English, K. (2013). Benchmark analysis for Pacific salmon conservation units in the Skeena watershed. A report to Pacific Salmon Foundation.
- Liller, Z. W., Hamazaki, H., Decossas, G., Bechtol, W., Catalano, M., and Smith, N. (2018). Kuskokwim River Chinook salmon run reconstruction model revision – executive summary.

Regional Information Report 3A.18-04, Alaska Department of Fish and Game, Anchorage, AK.

Ludwig, D. and Walters, C. J. (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):711–720.

Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J., Besbeas, P., Gimenez, O., and Thomas, L. (2014). *Modelling Population Dynamics*. Springer New York.

Plummer, M. (2017). *JAGS Version 4.3.0 User Manual*.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Research Board of Canada*, 11(5):559–623.

Rose, K. A., Cowan, J. H., Winemiller, K. O., Myers, R. A., and Hilborn, R. (2001). Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. *Fish and Fisheries*, 2(4):293–327.

Schindler, D. E., Armstrong, J. B., and Reed, T. E. (2015). The portfolio concept in ecology and evolution. *Frontiers in Ecology and the Environment*, 13(5):257–263.

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., and Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, 465(7298):609–612.

Schnute, J. T. and Kronlund, A. R. (2002). Estimating salmon stock <u+0096>recruitment relationships from catch and escapement data. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(3):433–449.

Smith, N. J. and Liller, Z. W. (2017). Inriver abundance and migration characteristics of Kuskokwim River Chinook salmon, 2015. Fishery Data Series 17-22, Alaska Department of Fish and Game, Anchorage, AK.

Staton, B. (2018). *codaTools: Utilities for Dealing with mcmc.lists*. R package version 0.1.1 available at <https://github.com/bstaton1/codaTools>.

Staton, B. A., Catalano, M. J., and Fleischman, S. J. (2017). From sequential to integrated bayesian analyses: Exploring the continuum with a pacific salmon spawner-recruit model. *Fisheries Research*, 186:237–247.

Su, Z. and Peterman, R. M. (2012). Performance of a bayesian state-space model of semelparous species for stock-recruitment data subject to measurement error. *Ecological Modelling*, 224(1):76–89.

Templin, W. D., Smith, C. T., and D. Molyneaux, L. W. S. (2014). Genetic diversity of Chinook salmon from the Kuskokwim River. USFWS Office of Subsistence Management,

Fisheries Resource Monitoring Program, Final Report 01-070, Alaska Department of Fish and Game, Anchorage, AK.

Walters, C. J. (1985). Bias in the estimation of functional relationships from time series data. *Canadian Journal of Fisheries and Aquatic Sciences*, 42(1):147–149.

Walters, C. J., Lichatowich, J. A., Peterman, R. M., and Reynolds, J. D. (2008). Report of the Skeena independent science review panel. A report to the Canadian Department of Fisheries and Oceans and the British Columbia Ministry of the Environment.

Walters, C. J. and Martell, S. J. D. (2004). *Fisheries Ecology and Management*. Princeton University Press. ISBN 978-0691115450.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.