

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 Simulation-based Evaluation of Assessment Approaches for Single-Species, Mixed-stock Pacific Salmon Fisheries | 2 |
| 1.1 Introduction | 2 |
| 1.2 Methods | 5 |
| 1.2.1 Regression-based models | 5 |
| 1.2.2 The full state-space model | 8 |
| 1.2.3 Alternate state-space models | 13 |
| 1.2.4 Analysis of Kuskokwim River substock data | 14 |
| 1.2.5 Simulation-estimation analysis | 16 |
| 1.2.6 Metrics of model performance | 17 |
| 1.3 Results | 17 |
| 1.4 Discussion | 17 |
| 2 Conclusions | 21 |
| Appendix A | 22 |
| Forecast Cross-Validation | 22 |
| Sliding Climate-Window | 24 |
| Bibliography | 25 |

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. 19

- 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. 20

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. | 18 |

Preface

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

Chapter 1

Simulation-based Evaluation of Assessment Approaches for Single-Species, Mixed-stock Pacific Salmon Fisheries

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. 2004), 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. 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, Schindler et al. 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 the substock 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, α). Stocks with higher α values can sustain greater exploitation rates than stocks with smaller

values, in fact, α can be expressed in terms of the exploitation rate that maximizes sustained yield (Schnute and Kronlund 2002):

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

Given that there is likely some level heterogeneity in α_j and $U_{\text{MSY},j}$ among individual substocks j , 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 of 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 perform 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 are often not precise enough to target particular substocks in the fishery, 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 (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 best-suited for Bayesian inference with Markov Chain Monte Carlo (MCMC) methods, but has been shown to reduce bias in estimates in some circumstances (Su and Peterman 2012, Walters and Martell 2004).

There has been recent interest in using multi-stock state-space spawner recruit models for policy analyses that incorporate notions of substock diversity as well as other fishery objectives (e.g., temporal stability of harvests). Before strong inferences can be made from such analyses, the performance of the estimation models used to parameterize them needs to be evaluated, as well as the appropriate level of model complexity. In this final chapter, I will evaluate the performance of a range of assessment models for mixed-stock salmon fisheries 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 using empirical data from Kuskokwim River Chinook salmon substocks, and
- (3) test the performance of the assessment models *via* simulation-estimation.

1.2 Methods

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_{MSY} and S_{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. I will begin by describing the estimation models assessed in this study and then provide details on the empirical and simulation analyses.

1.2.1 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) n_j independent regression models. A description and justification of each method is provided in the sections that follow.

1.2.1.1 Mixed-effect linear regression

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

$$R_y = \alpha S_y e^{-\beta S_y + \varepsilon_y} \quad (1.2)$$

where R_y is the total recruitment produced by the escapement S_y in brood year y , α is the maximum recruits-per-spawner (RPS), β is the inverse of the escapement that produces maximum recruitment (S_{MAX}), and ε_y are independent mean zero normal random variables attributed solely to environmental fluctuations. Primary interest lies in estimating the population dynamics parameters α and β 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:

$$\log(\text{RPS}_y) = \log(\alpha) - \beta S_y + \varepsilon_y, \quad (1.3)$$

allowing for estimation of the parameters $\log(\alpha)$ and β 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. 2002). Regression-based methods to 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 α and negative biases in β , causing the same directional biases in U_{MSY} and S_{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 (but false) scatter which inserts variability that commonly-used regression estimators do not account for (Walters and Ludwig 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 $[\log(\alpha)]$:

$$\log(\text{RPS}_{y,j}) = \log(\alpha_j) - \beta_j S_{y,j} + \varepsilon_y, \quad (1.4)$$

where

$$\log(\alpha_j) = \log(\alpha) + \varepsilon_{\alpha,j}, \quad (1.5)$$

and

$$\varepsilon_{\alpha,j} \sim N(0, \sigma_\alpha). \quad (1.6)$$

It does not make sense to include stock-level random effects on the slope, given that β 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.

1.2.1.2 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 α_j (those in the tails of the hyperdistribution) toward α . This behavior may not be preferable for policy

recommendations, as it should tend to dampen the extent of heterogeneity estimated in α_j . For this reason, independent regression estimates for each substock will also be obtained (i.e., the full fixed effects model) for evaluation.

1.2.1.3 Brood table reconstruction

An important point in the use of the regression-based method is in how $RPS_{y,j}$ is obtained. Only $S_{y,j}$ is directly observed; $R_{y,j}$ is observed (for Chinook salmon) over four 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 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 $y + 7$. Thus, it is easy to see how missing one year of sampling (which is an incredibly common occurrence, Figure 1.2) can lead to issues with this approach. Only completely observed $RPS_{y,j}$ observations will be used for this analysis, with the exception of missing age count data. For substocks with no age composition data, the average age composition across substocks that have data will be used to reconstruct $RPS_{y,j}$, but will be provided only for years with escapement sampling for substock j . Only substocks with ≥ 3 completely observed pairs of $S_{y,j}$ and $RPS_{y,j}$ were fitted.

1.2.2 The full state-space model

There will be four versions of the state-space formulation. As three versions are simplifications of the full model, the full model will be presented completely and the changes resulting in the other three model structures will be 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) escapement counts from the n_j substocks with data observed over n_t calendar years (some of which may be missing observations),
- (2) n_t calendar year estimates of aggregate harvest summed across all substocks included in the analysis, and
- (3) vectors of length n_a representing the calendar year age composition (relative contribution of each age class to the total run) for all substocks that have this information.

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.

1.2.2.1 Process submodel: brood year processes

The recruitment process operates by producing a mean prediction from a deterministic Ricker (1954) relationship (Equation (1.2)) for n_y brood years for each of the n_j substocks. From these deterministic predictions, auto-correlated 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 are not linked to a spawner abundance through Equation (1.2), but instead will be 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):

$$\bar{R}_{y,j} = \frac{\log(\alpha_j)}{\beta_j}, \quad (1.7)$$

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

$$\bar{R}_{y,j} = \alpha_j S_{t,j} e^{-\beta_j S_{t,j}}, \quad (1.8)$$

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

From these deterministic predictions of the biological recruitment process, auto-correlated lag-1 process errors will be added to produce the true realized states:

$$\log(R_{y,1:n_j}) \sim \text{MVN}(\log(\bar{R}_{y,1:n_j}) + \omega_{y,1:n_j}, \Sigma_R), \quad (1.9)$$

where

$$\omega_{y,1:n_j} = \phi(\log(R_{y-1,j}) - \log(\bar{R}_{y-1,j})), \quad (1.10)$$

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 auto-correlation, ϕ is the lag-1 auto-correlation coefficient, and Σ_R is a covariance matrix representing the white noise portion of the total recruitment process variance. The covariance matrix Σ_R will be estimated such that each substock will have a unique variance and covariance with each other substock. The multivariate normal errors are on the log scale so that the variability on $R_{y,j}$ is lognormal,

which is the most commonly used error distribution for describing spawner-recruit data sets (Walters and Martell 2004). Further, the multivariate normal will be 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 is 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 will be used here with maturity schedules shared perfectly among substocks within a brood year. Brood year-specific maturity schedules will be treated as random variables such that:

$$p_{y,a} \stackrel{\text{iid}}{\sim} \text{Dir}(\pi_{1:n_a} D). \quad (1.11)$$

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

1.2.2.2 Process submodel: calendar-year processes

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

$$N_{t,j} = \sum_{a=1}^{n_a} R_{t+n_a-a,j} p_{t+n_a-a,a}, \quad (1.12)$$

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:

$$H_{t,j} = N_{t,j} U_t v_j, \quad (1.13)$$

and escapement will be obtained as:

$$S_{t,j} = N_{t,j} (1 - U_t v_j), \quad (1.14)$$

where v_j are substock-specific vulnerabilities to harvest (1 = fully vulnerable; 0 = not vulnerable at all). For the analysis of empirical Kuskokwim River data, these quantities will be externally reconstructed by region using historical run and harvest timing. For the simulation analysis, all substocks will be assumed fully vulnerable for simplification. 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 will be 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 .

1.2.2.3 Observation submodel

Three data sources will 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 will be converted to lognormal standard deviations:

$$\sigma_{\log} = \sqrt{\log(\text{CV}^2 + 1)}, \quad (1.15)$$

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 will be fitted using parameter vectors $q_{t,1:n_a,j}$ and observed vectors of $(q_{obs,t,1:n_a,j} ESS_{t,j})$.

1.2.3 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 Σ_R will be constructed by estimating a single σ_R^2 and ρ , and populating the diagonal elements with σ_R^2 and off-diagonal elements with $\rho\sigma_R^2$. One drawback of constructing Σ_R this way is that $\rho < -0.05$ for a 13×13 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.

1.2.4 Analysis of Kuskokwim River substock data

1.2.4.1 Data sources

AYKDBMS

1.2.4.2 Data preparation

Need to turn partial aerial surveys into total annual escapement estimates

1.2.4.2.1 Substock escapement

For substocks monitored *via* weir, $S_{obs,t,j}$ was taken to be the total estimated weir passage each year (**CITE ADF&G REPORT**) with a CV of 5%. 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, resulting in them being indices of escapement rather than estimates of total escapement. The later of these two information sources was desired however, because it allows calculation of biological reference points that are expressed in terms of the scale of the population (e.g., S_{MAX}), rather than as a rate (i.e., U_t). Note that if only estimates of U_{MSY} were required, no accounting for the partial count would be necessary.

The approach developed to estimate total escapement from single-pass aerial surveys involved:

- (a) mapping the distribution of detected telemetry-tagged Chinook salmon against distribution of the aerial survey counts. This comparison allowed for an expansion to estimate how many salmon would have been counted if the entire tributary had been flown.
- (b) obtaining and applying an “observability” correction factor for the temporal 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.

Spatial expansion

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

$$\frac{A_{f,t,j}}{T_{f,t,j}} = \frac{A_{u,t,j}}{T_{u,t,j}}, \quad (1.16)$$

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 substock j . 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 (1.16) and can be rearranged as:

$$A_{u,t,j} = A_{f,t,j} \frac{T_{u,t,j}}{T_{f,t,j}}, \quad (1.17)$$

If we assume $T_{u,t,j}$ is a binomial random variable, we have:

$$T_{u,t,j} \sim \text{Binomial}(\pi_j, T_{u,t,j} + T_{f,t,j}). \quad (1.18)$$

Here, π_j represents the probability that a tagged fish in spawning tributary j was outside of the survey flight reach at the time of the aerial telemetry flight. If we rearrange π_j to be on the odds scale, we have:

$$\psi_j = \frac{\pi_j}{1 - \pi_j}. \quad (1.19)$$

Estimated expansion factors are shown in Table ???. The odds value ψ_j can be substituted for the division term in Equation (1.17) which gives:

$$A_{u,t,j} = A_{f,t,j} \psi_j. \quad (1.20)$$

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

$$\hat{A}_{t,j} = A_{f,t,j} + A_{u,t,j}. \quad (1.21)$$

Substitution of Equation (1.20) into (1.21) and factoring gives the estimator:

$$\hat{A}_{t,j} = A_{f,t,j}(1 + \psi_j). \quad (1.22)$$

1.2.4.2.2 Aggregate harvest

1.2.4.2.3 Age composition

1.2.5 Simulation-estimation analysis

1.2.5.1 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.

1.2.5.2 Operating model: Observation submodel

For a given set of simulated true states, a set of observed states ($S_{obs,t,j}$, $H_{obs,t}$, $q_{obs,t,a}$) 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 ($H_{obs,t}$) will be available every year in each simulation.

1.2.6 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 |

```
## Warning: package 'dplyr' was built under R version 3.5.1
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
## Warning: package 'kableExtra' was built under R version 3.5.1
```

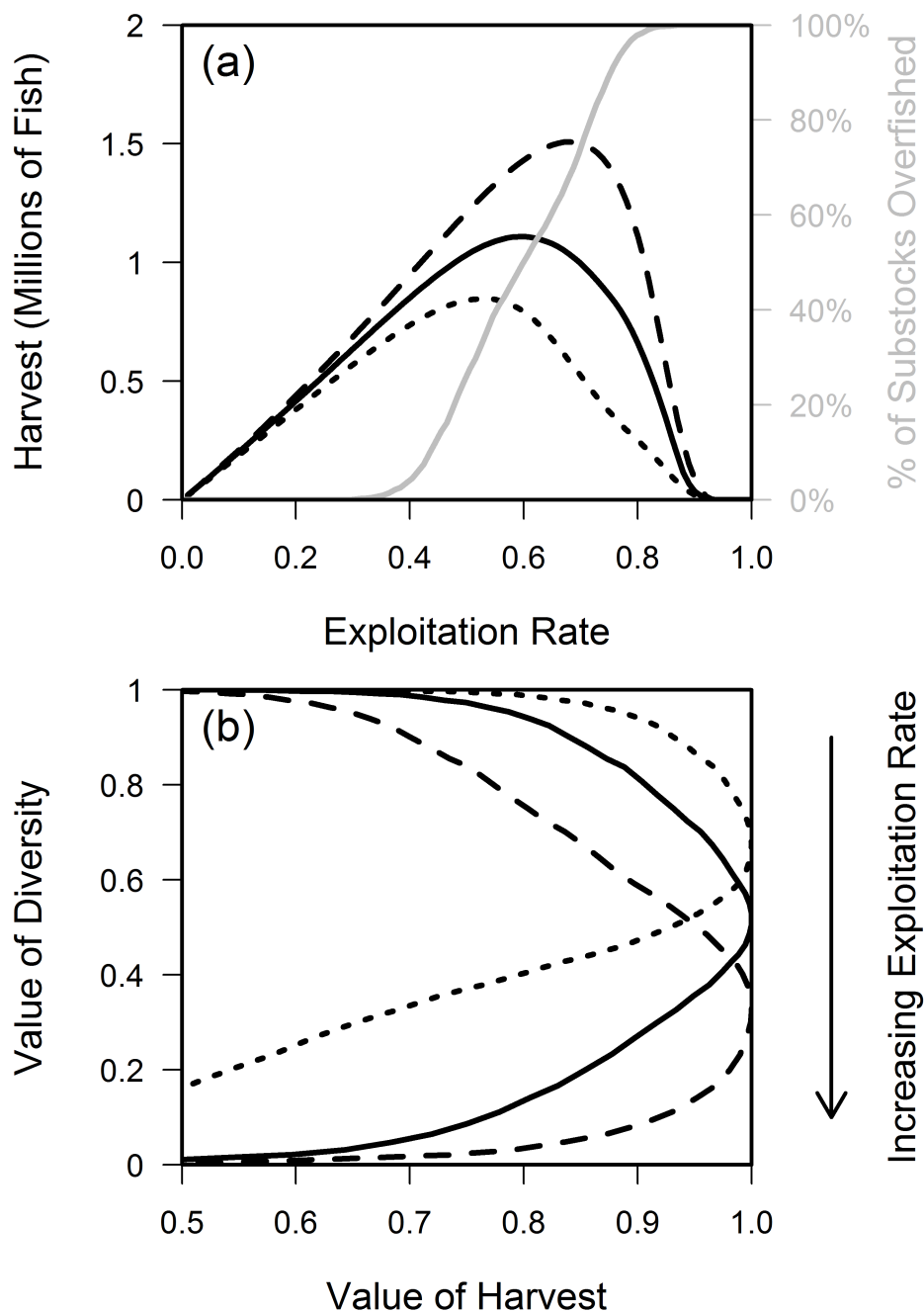
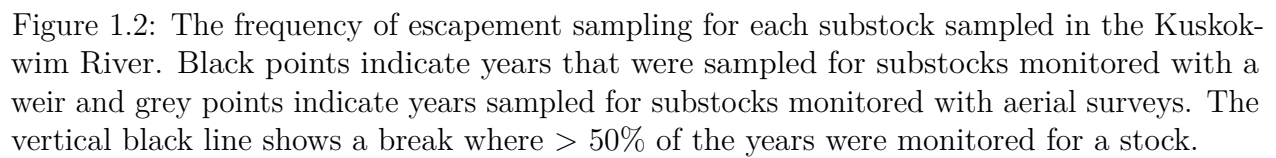



Figure 1.1: Visualization of how different types of heterogeneity in substock productivity and size influence the shape of trade-offs in mixed-stock salmon fisheries. Solid black lines are the case where stock types are split evenly among large/small and productive/unproductive stocks. Dotted black lines 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.



Chapter 2

Conclusions

This chapter contains my thoughts on the topic of the dissertation. What was found, what will be useful to use in the future, what should be looked at in more detail?

Appendix A

This appendix contains the necessary code to perform two of the main parts of the run timing forecast model approach in Chapter ??.

Forecast Cross-Validation

Function Name

`forecast.CV`

Purpose

Arguments

Arguments:

1. `x`: a vector containing the time series of the x-variable
2. `y`: a vector containing the time series of the y-variable
3. `start.ind`: the index to start the forecast cross-validation (e.g., 10 would train to 10 years and start forecasting in the 11th, then continue until present).
4. `na.rm`: logical; do you wish to remove NA observations before calculating summary statistics?
5. `include.last.year.in.scores`: logical; do you wish to have the last year of `y` to influence the cross-validation score?

Psuedocode

Source Code

```
forecast.CV = function(x, y, start.ind, na.rm = F, include.last.year.in.scores = T) {  
  # total number of observed pairs  
  n = length(x)
```

```

# validation end years
val.end = start.ind:(n-1)
n.val = length(val.end)
# containers
error = numeric(n.val)
abs.error = numeric(n.val)
pred.se = numeric(n.val)
# containers for training data
train.x = list()
train.y = list()
# containers for validation data
val.x = numeric(n.val)
val.y = numeric(n.val)
pred.val.y = numeric(n.val)
for (i in 1:n.val) {
  # indices to train and validate over
  train.ind = 1:val.end[i]
  val.ind = max(train.ind) + 1
  # store the training data
  train.x[[i]] = x[train.ind]
  train.y[[i]] = y[train.ind]
  # store the validation data
  val.x[i] = x[val.ind]
  val.y[i] = y[val.ind]
  # fit model to training data
  temp.x = train.x[[i]]; temp.y = train.y[[i]]
  fit = lm(temp.y ~ temp.x)
  sig = summary(fit)$sigma
  # forecast
  pred.val.y[i] = predict(fit, newdata = data.frame(temp.x = val.x[i]))
  # statistics
  error[i] = val.y[i] - pred.val.y[i]
  abs.error[i] = abs(error[i])
}

if (!include.last.year.in.scores) {
  error[n.val] = NA
  abs.error[n.val] = NA
}

# return output
output = list(error = error, abs.error = abs.error, mae = mean(abs.error, na.rm = na.rm))

```

```
    return(output)  
}
```

Sliding Climate-Window

Bibliography

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