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

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

List of Figures

List of Tables

Preface

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

Chapter 1

Introduction

Pacific salmon (*Oncorhynchus* spp.) constitute an integral natural resource in Alaska to subsistence, commercial, and recreational interests. There is a long history of resource development, exploitation, regulation, and dependence on these resources within the region (Cooley, 1963). In many cases, the resource use is dictated by the locality of the system; for example, stocks located near urban areas are often primarily exploited by recreational fishers whereas more remote stocks often constitute commercial and/or subsistence uses. This proposed dissertation focuses on the challenges and solutions in the assessment and management of these more remote stocks and the subsistence fisheries that rely heavily upon them.

Wild Pacific salmon represent a fantastic natural resource, which results largely from their unique life history strategy. Pacific salmon exhibit a migratory strategy known as anadromy: they spawn in freshwater where eggs hatch and juveniles rear for several years. Juveniles then migrate to the ocean where they spend the majority of their lives feeding on abundant prey resources. Once reaching maturity, adults return to their natal streams to spawn and complete the life cycle. The result of this life history strategy is an incredibly productive resource that grows entirely on its own and delivers itself to harvesters when the time comes for exploitation.

Like for all exploited natural resources, the management of Pacific salmon fisheries involves making decisions about how to exploit the resource in order to best attain a suite of biological, social, and economic objectives (Walters, 1986). These decisions are intrinsically difficult due to conflicting objectives and uncertainties in system state, system response to management actions, and implementation (Walters and Holling, 1990). Put another way, assuming a manager knows exactly what they wish to obtain, getting there is made difficult by not knowing how large the harvestable surplus is, how the stock will respond to harvesting, or that their management action will actually obtain what is desired. Despite these difficulties, a decision must be made (without decision-making there is no management; Hilborn and Walters, 1992) and the consequences, whether favorable or undesirable, must be accepted. Thus, I would argue that the science of monitoring, assessment, and prediction in the context of Pacific salmon fisheries is tasked with identifying trade-offs and reducing uncertainty, both of which insert difficulty into the decision-making process as discussed below.

The management of Pacific salmon fisheries can be thought of as a hierarchy of (1) guiding objectives, (2) management strategies to attain objectives, and (3) tactics to implement the management strategies (Table 1.1). At the upper level, long-term decisions are made about the objectives of the resource exploitation. These long-term objectives constitute what could be referred to as fundamental objectives: they are desired endpoints, but do not at all imply how they should be attained. These fundamental objectives often involve notions of sustainability and maintenance of biological diversity and often include social objectives such as maximization and stability of harvest. Already, it is clear that these fundamental objectives are often conflicting. For example, consider the objective of maximizing harvest: in fisheries that harvest multiple stocks (i.e., distinct spawning units), oftentimes maximum harvest may only obtained by overexploiting weak stock components and possibly reducing diversity. As another example, consider the objective of long-term sustainability: in order to ensure that the stock is sustained, some level of harvest fluctuations must be accepted (lower harvests must be allowed when the stock is at low abundance). These conflicting objectives imply that trade-offs exist (all objectives cannot be maximized simultaneously). It is worth noting here that the decisions made at the uppermost level of the management hierarchy are purely social and economic and salmon stock assessment scientists should play little-to-no advisory or advocacy roles in making these decisions (Walters and Martell, 2004). The Policy for the Management of Sustainable Salmon Fisheries (5 AAC 39.222) states that the objectives of salmon management in Alaska are

> "... to ensure conservation of salmon and salmon's required marine and aquatic habitats, protection of customary and traditional subsistence uses and other uses, and the sustained economic health of Alaska's fishing communities".

The policy goes on to say that managers should target "... to the extent possible, maximum sustained yield [MSY]".

The second level of the management hierarchy is made up of harvest strategies and policies that guide how the long term objectives are to be obtained. The State of Alaska has selected the fixed escapement policy as the management strategy to obtain the long-term objectives of sustainability and yields that are close to the maximum. These escapement goals are given as ranges that dictate the target number of spawning adults each year; any portion of the stock above the escapement goal is considered surplus (excess biological production) and should be harvested. Uncertainty at this intermediate level of the management hierarchy (i.e., regarding the optimal escapement goal) is often a result of incomplete understanding of system status and function. For example, in order to determine what the optimal escapement goal should be to obtain MSY, knowledge of stock productivity and carrying capacity are required. These quantities are often derived using spawner-recruit analyses, which are inherently uncertain: data are rarely informative about the shape of the true underlying spawner-recruit relationship but instead provide a probabilistic distribution of expected outcomes at a given stock size (Walters and Martell, 2004). Traditionally, it has been thought that these uncertainties can be reduced by more monitoring and the development of rigorous assessment and prediction models to better understand system function. However, it has often been argued that while monitoring and assessment models are obviously important (performance relative to objectives must be measured), true understanding of system behavior comes only from experimentation in management (the concept of "active adaptive management"; Walters, 1986). The classic example is to assess the maximum productivity of the stock (i.e., in the absence of density dependent mortality), the spawning stock must be forced to small sizes and the resulting recruitments must be observed. However, management actions that ensure these observations are made are undesirable to many managers and stakeholders, considering that exploiting a stock down to these low levels is risky (Walters, 1986).

At the lower level in the management hierarchy, intra-annual (or in-season) decisions are made regarding how to exploit the current year's run to attain the long-term objectives. In other words, given a management strategy (i.e., fixed escapement), the manager is still tasked with deciding how to best implement the fishery within a year to ensure the strategy is followed. As will be illustrated in this proposed dissertation, these decisions at the intra-annual level of the management hierarchy are often poorly informed by data which often results in indecisiveness, subjectivity, non-transparency, frustration, and missed opportunities.

This proposed dissertation will be partitioned into three chapters, each that expands on and investigates the performance of potential solutions to the aforementioned difficulties in Pacific management in Western Alaska. Each chapter will focus on the Kuskokwim River drainage in Western Alaska, which is characterized by being a large drainage (>50,000 km^2), harvests are taken by primarily subsistence users who are nearly all native Alaskans, and the primary species of interest being Chinook salmon. Although this proposed dissertation is quite narrow in its geographical and biological focus, a wide range of management issues will be addressed and the developed tools and assessment methods will be evaluated thoroughly. Furthermore, the concepts and tools discussed, developed, and evaluated will be generalizable to other stocks with similar spatial structures, exploitation characteristics, and population dynamics.

Chapter 2 will work at the intra-annual level of the hierarchy to develop and evaluate the performance of a run timing forecast model that can be used to aid in the interpretation of in-season data. As will be shown, uncertainty in run timing makes the interpretation of in-season data incredibly difficult. For example, consider a case in which an abundance index has produced high counts early in the season when compared to the historical average. This observation is consistent with at least two run scenarios: (1) a small run with early run timing or (2) a large run with average timing. There is a large discrepancy in the amount of harvestable surplus suggested by each of these discrete scenarios, leading to a large amount of uncertainty about how the fishery should be executed. The overall objective of Chapter 2 will be to develop and evaluate the reliability of a run timing forecast model for Kuskokwim River Chinook salmon. A secondary goal of Chapter 2 will be to formally assess the utility of having access to the run timing forecast model in terms of reducing uncertainty and bias in run size indices used in intra-annual management decisions.

Chapter 3 will again address the lowest level of the management hierarchy (i.e., intra-annual decision-making), but in this case in a more direct sense using an analysis framework known broadly as Management Strategy Evaluation (MSE). This analysis will seek to evaluate several harvest control tactics to identify strategies that perform well at attaining pre-defined objectives (e.g., meeting the escapement goal, distributing harvest equally across villages and stock components, etc.) across a range of biological states (e.g., run size, stock composition, and run timing). This analysis is needed because while the fixed escapement strategy seems simple to execute, actually doing so is made difficult largely due to uncertainty regarding the size of the incoming run (i.e., the amount of harvestable surplus is not known). Additionally, there may be a set of tactics that perform well at limiting harvest in low run size years but doing so in a "fair way', where the burdens of shortages aren't carried primarily by a certain subset of resource users. If a consistent set of rules or triggers could be identified that perform reasonably well at meeting management objectives without precise knowledge of run size, it could prove very useful to managers and decision making within the region.

Chapter 4 will move up the hierarchy to the second level and will seek to extend the single stock assessment models currently used in many systems in Alaska to multi-stock assessments. When an aggregate stock is made up of several distinct components, each with their own productivity, it is likely that exploitation at some level (e.g., 50%) results in the more productive components being under-exploited while the weaker stocks may be over-exploited. This reality implies a trade-off: to preserve stock diversity, some harvest must be foregone. Before the shape and magnitude of these types of "harvest-biodiversity" trade-offs are quantified, some understanding of the variation in substock productivity and carrying capacity is required. The multi-stock assessment framework developed in Chapter 4 will be tailored to provide this information for these sorts of trade-off analyses and others that require similar information sources. Multi-stock assessments may assume one of several different model structures (e.g., by fitting separate models to the data from each stock or by fitting a single model to all data simultaneously). In some cases, one approach may be preferable over the other, and a primary objective of Chapter 4 will be to evaluate the performance of a range of assessment strategies (in terms of accuracy and precision) across a range of biological and data quality conditions.

Table 1.1: One way of viewing the structure of renewable natural resource (including salmon) management as described in the text, including examples of alternatives and sources of uncertainty at each level.

Examples	**Sources of Uncertainty**
Overarching Objectives	
Ensure sustainability	Relative importance of objectives
Maximize harvest	Problem boundaries
Stabilize harvest	
Maximize economic value	
Inter-annual Strategies	
Constant escapement	Stock productivity
Constant exploitation rate	Stock status
Constant catch	Drivers of stock change
Adaptive exploitation	Shape/magnitude of trade-offs
Intra-annual Tactics	
Triggers and thresholds	Harvestable surplus
Time, area, gear restrictions	Uninformative data
Limited participation	Fisher behavior

Chapter 2

Development and Evaluation of a Migration Timing Forecast Model for Kuskokwim River Chinook Salmon

Abstract

Annual variation in adult salmon migration timing makes the interpretation of in-season assessment data difficult, leading to much in-season uncertainty in run size. We developed and evaluated a run timing forecast model for the Kuskokwim River Chinook salmon stock, located in western Alaska, intended to aid in reducing this source of uncertainty. An objective and adaptive approach (using model-averaging and a sliding window algorithm to select predictive time periods, both calibrated annually) was adopted to deal with multidimensional selection of four climatic variables and was based entirely on predictive performance. Forecast cross-validation was used to evaluate the performance of three forecasting approaches: the null (i.e., intercept only) model, the single model with the lowest mean absolute error, and a model-averaged forecast across 16 nested linear models. As of 2016, the null model had the lowest mean absolute error (2.64 days), although the model-averaged forecast performed as well or better than the null model in the majority of retrospective years. The model-averaged forecast had a consistent mean absolute error regardless of the type of year (i.e., average or extreme early/late) the forecast was made for, which was not true of the null model. The availability of the run timing forecast was not found to increase overall accuracy of in-season run assessments in relation to the null model, but was found to substantially increase the precision of these assessments, particularly early in the season.

2.1 Introduction

In-season management strategies for Pacific salmon (*Oncorhynchus* spp.) fisheries rely heavily on indices of in-river abundance (e.g., test fisheries, sonar counts, etc.) to inform harvest control rules that attempt to attain the balance of meeting pre-determined escapement objectives while allowing adequate opportunity for harvest (Catalano and Jones, 2014). However, because indices of abundance are confounded by the phenology (i.e., timing) of the migration, their interpretation is very difficult in-season. For example, smaller-than-average index values early in the season could be due to either a small run with average timing or by a late large run, when interpreted in the context of historical years (Adkison and Cunningham, 2015). This ultimately leads to great uncertainty about how much of the incoming run has passed, which is a key piece of information that dictates fishery harvest opportunities. There exists no information in the current year's abundance index to inform the manager if (for example) 25% or 75% of the run has passed on any given day. Yet, depending which is true, the optimal management decision could be vastly different. Thus, in-season assessment typically involves some characterization of the variation in historical run timing to formulate a range of possible run size scenarios that could be representative of the current year's run size. However, given the amount of variation in historical run timing, these scenarios are rarely informative during the majority of the migration, when key harvest decisions are being made because the run scenarios may span all possible run sizes. As a result, the pre-season run size forecast remains the most precise piece of information for much of the season. If it were possible to predict the timing of the incoming run (e.g., earlier- or later-than-average) with some level of confidence, it could prove valuable for in-season assessment and decision-making by reducing uncertainty in run size predictions.

While previous research has uncovered several key physiological mechanisms that are involved with natal homing (Hasler and Scholz, 1983) and return migrations of adult salmon to freshwater environments (Cooperman et al., 2010; Cooke et al., 2008; **?**), the exact physiological and behavioral responses of adult salmon to relatively small-scale environmental gradients within estuaries, which are likely the ultimate determinants of freshwater entry timing, are still poorly understood. Despite this uncertainty, several hypotheses have been put forth that are broadly consistent with the observed timing patterns of several species across a large geographic area (i.e., western and southwestern Alaska). Two primary influences have been suggested: genetic (Quinn et al., 2000; Anderson and Beer, 2009; O'Malley et al., 2010) and environmental (HODGSON et al., 2006; Keefer et al., 2008) mechanisms. Substantial evidence exists to suggest that both genetic and environmental controls are involved in determining migration timing, however it is broadly thought that genetic variation influences sub-stock variation (i.e., different tributary spawning groups within the same major river basin) and environmental variation influences the timing of the aggregate (i.e., basin-wide) run (Keefer et al., 2008; Anderson and Beer, 2009). This is consistent with the notion that genetically distinct components of the aggregate run behave differently as a result of their life history strategies and/or the characteristics of their specific spawning grounds (e.g., sub-stocks that must travel farther in-river to reach spawning grounds enter freshwater earlier; Clark et al., 2015; sub-stocks that spawn in tributaries influenced by warmer lakes enable later spawning; Burger et al., 1985) but that certain environmental conditions act on the aggregate run to either hasten or delay freshwater entry. It has also been suggested that run size may have an influence on migration timing, although empirical support for this claim seems to be lacking. If there were indeed relationships between run timing and run size, these need to be quantified as certain combinations are particularly troublesome for managers (e.g., small/early runs and large/late runs appear the same early in-season; Adkison and Cunningham, 2015).

At the aggregate population scale, which is the focus of this Chapter, it has been observed that migrations occurring in the spring and summer generally occur earlier in years with warmer spring temperatures [Mundy and Evenson (2011); hodgson-etal-2006]. Mundy and Evenson (2011) suggested that this pattern may be explained by the stability of the estuarine water column where adult salmon stage in preparation for riverine entry (or alternatively, marine exit). High estuarine water column stability was hypothesized to impede riverine entry through two mechanisms:

(1) by presenting an osmotic barrier between freshwater riverine discharge and the saline ocean water which prevents osmotically incompetent individuals from crossing and,
(2) by preventing freshwater competent individuals from receiving olfactory cues essential to the homeward migration.

Thus, Mundy and Evenson (2011) hypothesized that years in which the estuarine water column is stable over a longer period of time would be associated with later migration timing. Although water column stability is a difficult variable to measure over large spatial scales, several variables that are known to influence it are available at large scales via remote sensing (e.g., satellite observations). Such variables are sea ice cover which prevents wind-driven mixing, associated local temperature-related variables like land-based air temperature or sea surface temperature (SST), and broader scale indicators such as the Pacific Decadal Oscillation (PDO), an index of temperature anomalies in the northern Pacific Ocean. Observational studies across the North American range of Chinook salmon have found environmental-run timing correlations that are consistent with this hypothesis (HODGSON et al., 2006; Keefer et al., 2008; Mundy and Evenson, 2011). Even if the water column stability hypothesis is incorrect, observed patterns suggest that environmental variables may be useful in forecasting run timing with some level of accuracy and certainty.

Several efforts have been made at exploiting these environmental-run timing relationships to develop run timing forecast models for Pacific salmon migrations. Mundy and Evenson (2011) developed a model for Yukon River Chinook salmon (*O. tshawytscha*) that used air temperature, SST, and ice cover to predict the day at which the 15th and 50th percentiles of the run passed a test fishery index location. Their model predictions fit the observed data well (nearly always within seven days, usually within three days), although out-of-sample predictive ability was not presented. Keefer et al. (2008) developed a similar framework for Columbia River spring run Chinook salmon and found run timing relationships with river discharge, river temperature, and ocean condition indices (e.g., PDO). Their best model explained 49% of the variation in median run timing with variation in the environmental variables. Anderson and Beer (2009) continued this work on the Columbia River spring Chinook stock, but added genetic components to their analysis based on the arrival timing of precocious males. Their findings revealed that both environmental variables and changes in abundance of genetically distinct populations, which had their own distinct migration timing and affected overall run timing of the spring Chinook salmon run in the Columbia River. These advancements have shown that relationships between migration timing and environmental variables exist and may have utility for use in forecasting efforts.

The Kuskokwim River, located in western Alaska, is the second largest river system in the state and supports culturally and economically important Chinook salmon fisheries. Chinook salmon return beginning in late May and continue through early August, with the median date of passage occurring between June 14 and July 2. Fisheries within the region harvest salmon in-river during freshwater migrations using primarily drift gillnet gear. The Kuskokwim River salmon fishery has a distinct cultural importance: nearly all inhabitants are native Alaskans belonging to the Yup'ik group and take salmon for subsistence purposes (Linderman and Bergstrom, 2009). While commercial salmon fisheries operate within the river, these fishers often also participate in subsistence take and revenues from the sale of commercially-harvested salmon often contribute directly to participation in subsistence activities (Wolfe and Spaeder, 2009). To ensure long-term sustainable harvest, the Chinook salmon fishery is managed with a drainage-wide escapement goal derived from an age-structured state-space spawner-recruit analysis (Hamazaki et al., 2012; Staton et al., 2017). To meet these pre-determined escapement goals, in-season management strategies implement time, gear, and area closures based on limited and imprecise information regarding annual run size. The distant locations of the majority of escapement assessment projects makes direct measurement of escapement performance unavailable until late in the season. Thus, the primary sources of run size assessment information are (1) a pre-season run size forecast range (obtained as the previous year's run size estimate $\pm \sim 20\%$) and (2) an in-river drift gillnet test fishery operated in Bethel, AK which has been implemented using consistent methods since 1984. The interpretation of this test fishery index suffers from the same issue of being confounded by run timing described earlier, making management decisions difficult. Without precise in-season indicators of run size, managers must often choose to either trust a pre-season run size forecast for the majority of the season or somehow place weights on the various run timing hypotheses when interpreting in-season data. Both options could lead to the wrong interpretation of the actual run size, which could have serious consequences for the management of the fishery in a given year (i.e., the unwarranted opening or closing the fishery resulting in severe under- or over-escapement). No published run timing forecast models currently exist for Kuskokwim River Chinook salmon but given the potential utility of independent run timing estimates for interpretation of in-season data, the development and evaluation of such a model is needed. The necessity of more accurate and precise in-season perceptions of run size is particularly evident in years with anticipated low runs, such as in recent years (i.e., since 2010), as this may allow managers to more effectively guard against over-exploitation while still allowing for limited harvest opportunities to support the cultural and subsistence needs of the region.

In this chapter, I present an analysis that develops and evaluates the performance of a run timing forecast model for Kuskokwim River Chinook salmon. The objectives were to

(1) quantify historical run timing,

(2) develop a run timing forecast model using environmental variables selected based on out-of-sample predictive performance
(3) assess the utility of the forecasting model for improving predictions of end-of-season test fishery indices of run size,
(4) determine if there is a relationship between run size and run timing for the Kuskokwim River Chinook salmon stock.

2.2 Methods

2.2.1 Estimates of migration timing

In this analysis, the forecasted quantity that represented migration timing was the day at which 50% of the run passed an index location (hereafter, D_{50}). To inform this quantity for each year in the analysis, we used daily catch-per-unit-effort (CPUE) data from the Bethel Test Fishery (BTF) operated by the Alaska Department of Fish and Game (ADF&G), which spans 1984 – 2016. The raw data were daily CPUE beginning on June 1 and ending August 24 each year. The cumulative sum of these daily CPUE values within a year follows a sigmoidal pattern reflecting the shape of the incoming salmon run which is characterized by relatively few early migrants, a peak where the majority of the fish are running, and relatively few late migrants. To estimate the median day of passage as a continuous variable, a logistic model was fitted to the cumulative proportion of daily CPUE of the form:

$$p_{d,t} = \frac{1}{1 + e^{-h_t(d - D_{50,t})}}, \tag{2.1}$$

where $p_{d,t}$ is the predicted cumulative proportion on day-of-the-year (DOY) d in calendar year t, h_t is the parameter that controls the steepness of the curve (i.e., duration of the run), and $D_{50,t}$ is the day at which 50% of the total annual CPUE was caught in year t. Annual estimates of $D_{50,t}$ and h_t were obtained by fitting $p_{d,t}$ to observed daily cumulative proportion by minimizing the sum of squared deviations from the model prediction. Uncertainty in these parameter estimates was not further considered in the analysis as the uncertainty was negligible. Further, the use of the BTF daily CPUE values to infer the location and shape of year-specific logistic timing curves made the assumption that these data provided an accurate representation of daily run strength within a year (i.e., that the influence of weather conditions or harvest on sampling was negligible).

2.2.2 Environmental variables

Environmental variables to be assessed for forecasting performance were chosen based on three criteria:

(1) previously established association with salmon run timing,
(2) availability for the Kuskokwim River during the years for which BTF index observations exist (1984 – 2016), and
(3) availability for use in a pre-season forecast model (i.e., available no later than June 10th in the year for which the forecasted value would be used).

Based on these criteria, four environmental variables were chosen for analysis: SST, percent sea ice cover (SIC), PDO, and land-based air temperature taken in Bethel, AK.

2.2.2.1 PDO data

Data collected for the PDO variable came from one of several indices produced by the National Oceanic and Atmospheric Administration (NOAA) (Mantua et al., 2017)[1]. The index is produced by taking the first principal component of monthly SST anomalies in the northern Pacific Ocean, after removing any global trends due to any systematic change over time (Mantua et al., 2017). Thus, for each year of the data set, a single monthly value was available for PDO. Previous studies have found PDO values prior to the initiation of the run have predictive value for Chinook salmon populations (Beer, 2007; Keefer et al., 2008).

[1]PDO data: http://research.jisao.washington.edu/pdo/PDO.latest.txt

2.2.2.2 Bethel air temperature data

Air temperature data for Bethel, AK were accessed from the Alaska Climate Research Center[2]. These data were available as daily means for each day of each year in the 1984 – 2016 data set.

2.2.2.3 SST and SIC

SST and SIC data were accessed from the NOAA Optimum Interpolation SST V2 High Resolution Dataset (Reynolds et al., 2007)[3]. These data were available as daily means for any 0.25° by 0.25° latitude by longitude grid cell on the globe. To limit the search, only grid cells within Kuskokwim Bay were selected for analysis Figure 2.1 as that is the area that Chinook salmon bound for the Kuskokwim River likely aggregate prior to riverine entry. The area with grid cells ranged from 58.5° N to 60° N by 164.25° W to 162° W, which resulted in a total of 54 0.25° latitude by 0.25° longitude grid cells. For SST, four grid cells fell partially over land (resulting in 50 grid cells with daily data) and for SIC, five grid cells were partially over land (49 grid cells with daily data). "Empty" grid cells were excluded and the remaining grid cells were used for prediction. Previous analyses have used a simple average over a wide spatial area (e.g., Mundy and Evenson, 2011) to create a single value for SST or SIC each year. However, this is somewhat arbitrary and does not account for the possibility of certain areas having stronger timing signals than others or that the areas with stronger signals may change over time. Thus, the gridded spatial structure of these variables was retained and the treatment of this structure in the forecast analysis is discussed below in Section 2.2.5.

[2]Alaska air temperature data: http://akclimate.org/acis_data
[3]Global gridded SST and SIC: http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.highres.html

2.2.3 Forecast model

To produce a forecast of run timing, relationships between historical observed pairs of the environmental variables each year and $D_{50,t}$ must be quantified. The simple linear regression framework was used to obtain these historical relationships:

$$\begin{aligned} D_{50,t} = \beta_0 + \beta_j x_{t,j} +, ..., +\beta_n x_{t,J} + \varepsilon_t, \\ \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma) \end{aligned} \tag{2.2}$$

where $D_{50,t}$ is the observed run timing value in year t, $x_{t,j}$ is the observed value of covariate j (of which there are J included in the model), β_0 and β_j are coefficients linking the observed values of $D_{50,t}$ with $x_{t,j}$, ε_t are random residual effects that explain deviations of observed $D_{50,t}$ from the fitted value and have constant variance equal to σ^2.

There are many such regression models that could be used to produce a run timing forecast (i.e., $\hat{D}_{50,t+1}$). This is because:

(1) there are four variables (PDO, air temperature, SST, and SIC) that could be included,
(2) each variable is temporally structured, i.e., there are daily or monthly values for each variable, and
(3) two variables (SST and SIC) are spatially-explicit, i.e., there are different values for each day and year for different areas of Kuskokwim Bay (Figure 2.1).

Item (1) deals with the specific values of j and J whereas items (2) and (3) deal with what values of x_t for a given j should take on.

2.2.4 Selection of predictive time periods

To fit the regression model in Equation (2.2), a single value for each $x_{t,j}$ was required. The covariate data were temporally structured, however, indicating that some selection of which time periods to use to populate $x_{t,j}$ was required. Oftentimes the average over an arbitrary time period, such as daily values in the month of February, is used based on *a priori* assumptions of the behavior of important factors (van de Pol et al., 2016). While this approach is simple to implement and explain, it is possible that a better time window (i.e., reliably more accurate) exists but was not considered. Furthermore, the importance of various time windows may change over time and the arbitrary selection of a single window does not allow for such changes to be detected. To avoid these issues, a rigorous temporal selection process, known as the sliding climate window algorithm (SCWA; van de Pol et al., 2016), was implemented to determine the best predictive time period for each variable considered in the forecast model. To find the most reliable temporal window for prediction, the SCWA evaluates all possible windows (subject to certain restrictions) over which to average for use as the predictor variable in the forecast model. The following section provides the details of the SCWA.

2.2.4.1 The SCWA

A "window" in this context is hereafter defined as a block of consecutive days in some portion of the year with starting day-of-the-year (DOY) denoted by D_F and ending day equal to D_L. The daily values within each evaluated window were averaged for the $x_{t,j}$ value to be used in a linear regression framework. As input constraints, the SCWA used in this analysis required:

(1) the start date of the first window to be evaluated (D_0),
(2) the end date of the last window to be evaluated (D_n), and
(3) the minimum window size of a candidate window ($\Delta_{D,min}$).

The algorithm started with the earliest and smallest possible time window: $D_F = D_0 = 1$ through $D_L = D_0 + \Delta_{D,min} - 1 = 5$. The performance of this window when used to obtain $x_{t,j}$ was evaluated (see Section 2.2.4.2 below) and the result was stored for comparison to other candidate windows. For the next window, D_F would remain at D_0, but D_L would be incremented by 1 day ($\ell = 1$). Thus, the endpoints of all candidate windows with $D_F = D_0$ can be generalized as:

$$[D_0, D_0 + \Delta_{D,min} - 1 + \ell], \tag{2.3}$$

for each $\ell = 0, 1, ..., n - 1$, where $n = D_L - D_F + 1$. For all windows, including those with $D_F = D_0$, this generalizes to:

$$[D_0 + f, D_0 + f + \Delta_{D,min} - 1 + \ell], \tag{2.4}$$

for each $f = 0, 1, ..., n - \Delta_{D,min}$ and $\ell = 0, 1, ..., n - \Delta_{D,min} - f$. Windows with $f > n - \Delta_{D,min}$ would contain fewer than $\Delta_{D,min}$ days and are thus prohibited. After evaluating all windows, the single window with the best predictive performance was used to obtain the forecast predictor variable for that data source (i.e., PDO *versus* air temperature). As an example, consider the following inputs:

- $D_0 = 1$ (i.e., the first day of the year),
- $D_n = 31$ (i.e., January 31), and
- $\Delta_{D,min} = 5$.

The SCWA would start with January 1 - January 5, then do January 1-6, January 1-7, etc., January 1-31. Next, it would exclude January 1 from consideration and evaluate all windows starting with January 2. When it completes the one window starting with January 27, it must stop because windows starting later than January 27 would result in windows shorter than 5 days. Example R code for how the sliding window algorithm was implemented is provided in Appendix A.

The values of D_0 and D_L for the four covariates are shown in Table 2.2. For air temperature, SST, and SIC $\Delta_{D,min}$ was set to 5. Note that because PDO was available in monthly values only, each month was treated as "day" in the algorithm described above and $\Delta_{D,min}$ was set to 1.

2.2.4.2 Forecast cross-validation

A metric was needed to measure the performance of the many windows. This metric was obtained using a time series forecast cross-validation procedure, which is an out-of-sample technique for data that are collected through time (Arlot and Celisse, 2010). The procedure operated by producing a forecasted value of D_{50} for year $t+1$ trained based on all data $x_{t,j}$ available from years $1, ..., t$. It then continued for all $t = m, ..., n-1$, where m is the minimum number of years necessary to fit the model (set at $m = 10$ in all cases) and n is the number of years of available data. Then, absolute forecast error in was calculated based on all forecasted years as $|D_{50,t+1} - \hat{D}_{50,t+1}|$, and yearly forecast errors were averaged to obtain mean absolute error ($\overline{\mathrm{AE}}$) which was used as the measure of model performance in window selection. The window with the lowest ($\overline{\mathrm{AE}}$) was selected as the optimal window to average over for prediction. The forecasting cross-validation procedure was used as opposed to other out-of-sample validation procedures, such as k-fold or leave-one-out methods, because the data were collected through time and the forecast model would never need to predict (for example) year 2010 from years 1984 – 2009 and 2011 – 2016, but rather it would always need to predict year $t+1$ from all previously-collected data. Example R code for how the forecast cross-validation was conducted is provided in Appendix A.

When forecasting $D_{50,t+1}$ from training data from $1, ..., t$, a single optimal climate window was selected for each variable and that window was used to estimate coefficients based on training data and obtain the environmental variable value for prediction in year $t+1$ to forecast $D_{50,t+1}$. When a new year of data was added to the training data (such as in the retrospective forecast analysis; Section 2.2.8), the optimal window for each variable was re-assessed using the algorithm again. For PDO and Bethel air temperature, which had no spatial structure, the SCWA was used to select the range of monthly (PDO) or daily (Bethel air temperature) values to include in the predictive climate window for each year in the analysis. For SST and SIC which contained a series of 50 and 49 grid cells, respectively, each with unique daily values, the SCWA was used on each grid cell separately. The result was 50 unique grid cell-specific windows for SST and 49 windows for SIC for each year of the analysis. The treatment of this spatial structure in the forecast analysis is discussed below in Section 2.2.5.

2.2.5 Evaluated forecast models

Linear regression [Equation (2.2)] was used to assess the forecast performance of each of the variables described above, both in isolation of and in combination with other variables. All possible subsets were evaluated (excluding interactive effects) for predictive ability through time, resulting in a total of 16 models ranging from the null (i.e., intercept only) model to the full (i.e., global) model (all four variables as additive predictors).

For the spatially-explicit variables (i.e., SST and SIC), a more complex treatment was required to prevent all grid cell values from being used as predictors in a single model. To handle the spatial structure, grid cell-specific regression models were fitted, then model-averaging based on AIC was used to obtain a single forecast D_{50} for each year (Burnham and Anderson, 2004). Under this approach, each grid cell g received an $\mathrm{AIC_c}$ score:

$$\mathrm{AIC}_{c,g} = n \log\left(\hat{\sigma}_g^2\right) + 2K + \frac{2K(K+1)}{n-K-1}, \tag{2.5}$$

where n is the number of data points used in each model, $\hat{\sigma}_g$ is the estimate of the residual standard deviation under grid g, and K is the number of model parameters. The corrected version of AIC ($\mathrm{AIC_c}$) is recommended in cases where the ratio of n to K is small (Burnham and Anderson, 2002). Then, each grid cell received a $\Delta\text{AIC}_\text{c}$ score, representing its relative performance in comparison to the best grid cell:

$$\Delta_g = \text{AIC}_{\text{c},g} - \text{AIC}_{\text{c},min}, \tag{2.6}$$

where $\text{AIC}_{\text{c},min}$ is the minimum AIC_c across all grid cells. Model (grid cell) weights were then calculated as:

$$w_g = \frac{e^{-0.5\Delta_g}}{\sum_j^G e^{-0.5\Delta_j}}, \tag{2.7}$$

where G is the number of grid cells. Grid cell-averaged predictions were then obtained as:

$$\hat{y}_{t+1} = \sum_g^G w_g \hat{y}_{g,t+1}, \tag{2.8}$$

where $\hat{y}_{g,t+1}$ is the forecasted value of D_{50} for grid cell g.

2.2.6 Forecast uncertainty

In addition to forecast accuracy, forecast precision is also of great importance. For models that did not require AIC_c model-averaging across grid cells, the following equation was used to produce a forecast standard error (SE):

$$\text{SE} = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_i^n (x_i - \bar{x})^2}}, \tag{2.9}$$

where n is the number of years the model was fitted to, x is the value of the predictor variable used for forecasting, and $\bar{x}$ is the mean of all predictor values excluding the new value used for forecasting. For models that used AIC_c model-averaging (i.e., those including SST and SIC), the following equation was used to produce prediction SE:

$$\text{SE} = \sum_{g}^{G} w_g \sqrt{\text{SE}_g^2 + (\hat{y}_{g,t+1} - \hat{y}_{t+1})^2}, \tag{2.10}$$

where SE_g is the prediction SE from grid cell g calculated using equation (2.9). This estimator of unconditional sampling standard error accounts for uncertainty within each model and the uncertainty due to model selection (Burnham and Anderson, 2004). Prediction intervals were calculated using the point estimate of prediction, the prediction SE, and appropriate quantiles from the corresponding t distribution.

2.2.7 Forecast model selection

Given 16 forecast models, it is impossible to know which will perform the best at forecasting for the current year. Thus, three methods to obtain a forecast for D_{50} were evaluated:

(1) the null (i.e., intercept only) model,
(2) the single model with the lowest forecast cross-validation score as of the last year, and
(3) model-averaging across the ensemble of 16 forecast models based on AIC_c scores.

According to Burnham and Anderson (2004), model-averaging should perform better than a single "best model" at prediction when there is a high degree of uncertainty about which model is best. This procedure was performed using equations (2.5) - (2.10), by substituting the prediction, prediction SE, and K for forecast model i, in place of grid g. Prediction intervals based on model-averaged predictions and prediction SE present somewhat of a problem when the different models contributing to the average contain differing degrees of freedom as it is unclear how many standard errors the prediction limits should lie from the mean prediction. Thus, the estimator suggested by Burnham and Anderson (2004) of the "adjusted SE" (ASE) was used:

$$ASE = \sum_{i}^{16} w_i \sqrt{\left(\frac{t_{df_i,1-\alpha/2}}{z_{1-\alpha/2}}\right)^2 \text{SE}_i^2 + (\hat{y}_{i,t+1} - \hat{y}_{t+1})^2}, \tag{2.11}$$

where $t_{df,i,1-\alpha/2}$ is the $1-\alpha/2$ quantile of the t distribution with degrees of freedom equal to that of model i and $z_{1-\alpha/2}$ is the corresponding quantile of the z (i.e., standard normal) distribution. The level $\alpha = 0.05$ was used in all cases.

2.2.8 Retrospective forecast analysis

The analysis was conducted in a retrospective forecast framework starting in 1994. All data after 1994 were ignored, optimal windows were selected for each of the four variables (and all grids for SST and SIC), all 16 models were fitted, a D_{50} forecast was made for 1995 using the three approaches described in Section 2.2.7, and each was evaluated for predictive accuracy. This process was repeated annually until the present (i.e., out-of-sample predictions made for 1995 – 2016), which allowed for the calculation of $\overline{\mathrm{AE}}$ through time as if the forecast model would have been available beginning in spring 1995. In addition to $\overline{\mathrm{AE}}$, median absolute error $\widetilde{\mathrm{AE}}$) was calculated to validate prediction accuracy of estimates by ignoring the effect of outlying poor predictions.

2.2.9 Value of forecast to run size assessments

It is important to remember that the purpose of producing a run timing forecast is to aid in the interpretation of in-season indices of run size such as test fisheries. To evaluate the utility of having access to the run timing forecast model, the accuracy and precision of an imperfect abundance index for the Kuskokwim River were compared when informed using D_{50} forecasts from the model-averaged and the null forecast models. The abundance index is denoted by EOS_t, and is the end-of-season cumulative CPUE observed in the BTF in year t. Under the assumption of constant catchability, EOS_t should be proportional to total abundance, with deviations introduced by sampling noise. In-season predictions of EOS_t were made for each year t, model i, and day d in the season with:

$$\widehat{\mathrm{EOS}}_{d,t,i} = \frac{\mathrm{CCPUE}_{d,t}}{\hat{p}_{d,t,i}}, \tag{2.12}$$

where $\mathrm{CCPUE}_{d,t}$ is the cumulative CPUE caught at the BTF through day d in forecasting year t $\left(\mathrm{CCPUE}_d = \sum_{j=1}^{d} CPUE_j\right)$, $\hat{p}_{d,t,i}$ is the predicted cumulative proportion of the run that had passed the BTF location on day d in year t from model i (i.e., model-averaged *versus* null forecast model) obtained by inserting the forecasted value of D_{50} into the logistic function [Equation (2.1)]. Uncertainty in the run timing forecast model was propagated to $\mathrm{EOS}_{d,t,i}$ using the delta method (Bolker, 2008). As the parameter h is also unknown in-season, the mean of all historical h_t was used as the point estimate, and the variance and correlation with $D_{50,t}$ was included in the covariance matrix supplied to the delta method. Accuracy was assessed using proportional bias $\left(\widehat{\mathrm{EOS}}_{d,t,i} - \mathrm{EOS}_t\right)/\mathrm{EOS}_t$ and precision was assessed using the SE of $\widehat{\mathrm{EOS}}_{d,t,i}$. Both accuracy and precision measures were compared between the null and the model-averaged forecast model on June 15, June 30, July 15, and July 30 each year a forecast was available (1995 - 2016). Using the null model to obtain $\hat{p}_d$ is essentially what would be done in the absence of an environmental forecast variable model for D_{50}.

2.2.10 Investigation of a run timing versus run size relationship

To test the hypothesis that run timing is related to run size (e.g., small runs are typically early, or *vice versa*), two models were investigated for their predictive performance using the forecast cross-validation criteria: the null model and a model that included run size as a predictive covariate in place of the environmental variables. Run size was obtained from a maximum likelihood run reconstruction model that compiles all assessment information (i.e., 20 escapement count indices, harvest estimates, drainage-wide mark-recapture estimates, etc.) to estimate the run size that makes the collected data most likely to have been observed

(Bue et al., 2012). The forecast absolute errors in each year were then compared using a two-tailed paired t-test using $\alpha = 0.05$.

2.3 Results

2.3.1 Estimates of run timing

The logistic curve fit the daily cumulative CPUE proportions well in all years of the BTF data set (Table 2.1), as indicated by an average residual standard error estimate of 0.022, with a maximum estimate of 0.038 in 1992. The majority (95%) of all residuals from all years fell between -0.056 and 0.044. Parameter estimates were quite precise, with $D_5 0$ having a smaller average coefficient of variation (CV) than h, (0.07% and 2.09%, respectively). Given this small degree of parameter uncertainty, it was ignored throughout the rest of the analysis.

2.3.2 Variable-specific relationships

Looking at each of the environmental variables in isolation of all others, it is clear that there is a distinct relationship between temperature-related environmental variables and Kuskokwim River Chinook salmon migration timing (2.2). For illustration purposes, the figures for the two gridded variables (SST and SIC) were produced by taking an average across all grid cells weighted by the AIC_c weight for each grid cell. Air temperature, PDO, and SST all had negative relationships with D_{50}, whereas SIC had a positive relationship (2.3).

2.3.3 Selected climate windows

It was difficult to generalize on the climate windows selected for each variable based on forecast cross-validation performance, because the selected windows changed with each new year of data and SST and SIC had windows for each grid cell; however, some noteworthy patterns arose. First, the best window for PDO was consistently the value for the month of May for each year the forecasts were produced (not shown). Second, selected windows for air temperature fluctuated from year to year to some extent, but long windows were consistently selected for all years after 1999, and generally spanned early February to late May (Figure 2.3a). Third, selected windows through time were substantially more variable for most grid cells for SST and SIC, although many grid cells remained relatively constant or became more "focused" as more years of data were added (Figures 2.3b1-4, 2.3c1-4). In general, chosen windows for SST began in early to mid-May and ended in late May or early June (Figure 2.3b1-4) whereas windows starting in early April and ending in mid to late April were predominately chosen for SIC (Figure 2.3c1-4). The selected climate windows in southern-most grid cells appeared more stable for SST (Figure 2.3 panels b3 and b4), whereas climate windows in northern grid cells appeared more stable for SIC (Figure 2.3 panels c1 and c2; stable in the sense that the optimal windows changed less as new years were added to the training data).

2.3.4 Forecast performance

Of the three investigated forecasting methods (null model, model with lowest forecast cross-validation error up to the forecasting year, and $\mathrm{AIC_c}$ model-averaging), the null model had the lowest $\overline{\mathrm{AE}}$ from 1995 to 2016 (2.64 days; Figure 2.4). $\mathrm{AIC_c}$ model-averaging performed better than using the single model with the lowest cross-validation score (3.04 versus 3.34, respectively; Figure 2.4). However, these patterns were not consistent across the entire time series. For the period of 1996 to 2008, the model-averaged forecast had a lower $\overline{\mathrm{AE}}$ than the null model, and for the period of 2009 to 2015, the model-averaged forecast had approximately the same or lower $\overline{\mathrm{AE}}$ scores (Figure 2.5). It was due in a large part to 2016 that the model-averaged forecast had a slightly higher $\overline{\mathrm{AE}}$ than the null model. Each model in the ensemble of 16 models (except the null) predicted an extremely early run in 2016 when in fact the observed run timing in 2016 was close to the historical (1984 – 2015) average (Figure 4). A similar case happened in 2015 (Figure 2.4). Expressing prediction error in terms of median absolute error ($\widetilde{\text{AE}}$) resulted in lower average errors (null = 2.08, single best = 2.56, and model-averaged = 2.35), indicating that extreme prediction errors (i.e., outliers) influenced the value of $\overline{\text{AE}}$. However, the relative differences in prediction error between models were approximately the same as for $\overline{\text{AE}}$, indicating outliers affected the prediction error scores for each model similarly. Additionally, by comparing the width of the prediction intervals in Figure 2.4 across forecasting approaches, it was clear that model-averaging substantially reduced prediction uncertainty (SE) in relation to the null and single best model approaches.

To compare performance in average versus extreme years among forecasting approaches, $\overline{\text{AE}}$ was further calculated in a more specific way: based on how similar or dissimilar the included years were to the mean observed run timing across all years. As would be expected, the null model performed well when only years with D_{50} within ± 1 days of the average were in included in the calculation of $\overline{\text{AE}}$ (Figure 2.6a), but its accuracy became increasingly worse as years with more extreme realized D_{50} values were included in the calculation (increasing x-axis values in Figure 2.6a). The two environmental variable forecast approaches (model-averaging or the single "best" model in each year) performed relatively equally well across this continuum and neither $\overline{\text{AE}}$ score was sensitive to the overall similarity or dissimilarity the included years had with average run timing (Figure 2.6a). The lower panel shows the relative frequency with which these various X scenarios occurred, indicating how much information each scenario contributed to the overall $\overline{\text{AE}}$. On the other hand, the null model only performed as well as the model-averaged forecast when years with $D_{50} \pm 0.5$ days outside of the mean were considered (Figure 2.6b). As only more extreme years were considered in $\overline{\text{AE}}$ (increasing x-axis values in Figure 2.6b), the null model rapidly performed worse and the model-averaged forecast remained relatively insensitive to the degree of extremity (Figure 2.6b).

2.3.5 Value to in-season run size assessments

When the model-averaged and null model forecasts for $D_{50,t}$ were retrospectively used to aid in in-season run assessment based on daily cumulative BTF CPUE, it was evident that the range of possible $\widehat{\text{EOS}}_{d,t}$ was substantially smaller when the model-averaged forecast was used as opposed to the null forecast. This is evident by the average daily standard errors of $\widehat{\text{EOS}}_{d,t}$ on June 15, June 30, July 15, and July 30: 256.97, 62.23, 6.69, 0.59 for the model-averaged forecast model and 337.19, 72.63, 6.90, and 0.57 for the null forecast, respectively. The reduction in the first two evaluated days is of key importance. In terms of accuracy, however, the forecast model did not perform better at informing EOS predictions than the null model. On the same days, the average proportional bias [(estimate – observed)/observed] using the model-averaged $D_{50,t}$ forecast was 0.152, 0.006, -0.024, -0.008 as opposed to 0.145, -0.035, -0.027, and -0.008 under the null model forecast. A visual example of two years is provided in Figure 2.7. The upper panels show the time series of $\widehat{\text{EOS}}_d$ when the model-averaged and the null model forecast models were used to inform the location of the logistic cumulative timing curve in 2013 and 2014. The horizontal line shows the observed EOS. 2013 is an example of when the null model would have been preferable to use (in terms of accuracy) and 2014 shows a case when the model-averaged forecast would have performed better.

2.3.6 Run timing versus run size relationship

There appeared to be no evidence to lend support for the hypothesis that run timing and run size are related for the Kuskokwim River Chinook salmon stock. On average, D_{50} occurred 0.01 (95% CL; -0.03 – 0.004) days earlier for each 1,000 fish increase in run size, which was not significantly different than no effect of run size on run timing ($p = 0.153, R^2 = 0.07$). Additionally, based on forecast cross-validation, the model that included run size did not perform better at prediction than the null model. On average, the null model resulted in an estimated absolute forecast error of 0.2 (95% CL; -1.07 – 0.68) days less than the run size model ($p = 0.64$).

2.4 Discussion

The environmental relationships with run timing I detected for the Kuskokwim River Chinook salmon stock are consistent with patterns found elsewhere in the region (e.g., Mundy and Evenson, 2011; HODGSON et al., 2006). Specifically, I found that warmer years were typically associated with earlier-than-average runs as were years with less-than-average SIC. These findings are consistent with the water column stability hypothesis suggested by Mundy and Evenson (2011). The amount of unexplained variation in the Kuskokwim model appears to be comparable between the Yukon River Chinook salmon stock as well (Mundy and Evenson, 2011). Using the relationships shown in Figure 2.2, the correlation with D_{50} was -0.52, -0.64, and 0.62 for air temperature, SST, and SIC, respectively. For the Yukon River Chinook salmon stock, Mundy and Evenson (2011) found correlations of -0.59, -0.72, and 0.66 for the same variables but measured at different spatial and temporal scales and with approximately 10 more years of data included (Table 2 in Mundy and Evenson, 2011). These similar correlations indicate the signals given by environmental variables are of relatively equal strength between these two systems. Given these similarities, future research applying our framework to Yukon River Chinook salmon (and other stocks) may enhance the applicability of our approach, if it could be shown that doing so would improve the accuracy of run timing predictions over other methods.

Given the overall strength of the environmental relationships, it is somewhat surprising that the null model forecast performed better on average than did the model-averaged forecast. This could, potentially, be due to the fact that a variety of biological (size; e.g., Bromaghin, 2005, and morphology; **?**) and abiotic factors (temperature; e.g., Salinger and Anderson, 2006, river discharge; e.g., Keefer et al., 2004, and migration distance; e.g., Eiler et al.,

2015) may affect migration rate (and subsequently, encounter probability) and catchability, introducing additional variability in my run timing estimates. Future research that accounts for these effects on encounter probability or catchability could offer improved predictions of run timing. Regardless of the underlying drivers, the overall prevalence of years with average run timing likely led to the enhanced performance of the null model. It is not surprising, however, that the model-averaged forecast performed better than the supposedly single "best" model. This finding is consistent with the literature on model-averaging predictions (**?**). Although the null model performed better in the long-term average (i.e., lower $\overline{\mathrm{AE}}$ as of 2016), there are reasons a manager may still justifiably prefer the model-averaged forecast. First, the difference in $\overline{\mathrm{AE}}$ between the model-averaged forecast and the null model was 0.4 days, which is small relative to the amount of annual variation in run timing (a 17 day range for D_{50} over 33 years). Second, the model-averaged forecast performed equally well in terms of forecast accuracy regardless of the type of run timing it was used to forecast (i.e., prediction error equal in extreme early/late and average years; Figure 2.6b). In contrast, the null model only performed comparably well in years with run timing within ± 3 days from average and error increased precipitously in more extreme years. Third, the 95% prediction intervals from the null model seemed too wide as 100% of the observations fell within the intervals, whereas 92% of the observations fell within the prediction intervals from the model-averaged forecast (which is closer to the ideal coverage, i.e., 95%). Prediction uncertainty was lower under the model-averaged forecast than the null model, which would ultimately lead to fewer run timing scenarios being considered to explain the observed in-season data (e.g., the earliest or latest scenarios could be excluded earlier in the season) leading to more certain interpretation of in-season indices of run size.

Table 2.1: Parameter estimates (mean with standard error in parentheses) from logistic curves from Equation 2.1 fitted to all years. $D_{50,t}$ is expressed as the day-of-the-year, for reference, day 174 is June 22$^{\text{nd}}$ in a leap year and June 21$^{\text{st}}$ in a non-leap year.

Year	$D_{50,t}$	h_t
1984	176.27	0.07
1985	169.38	0.02
1986	168.87	0.08
1987	173.39	0.05
1988	169.21	0.03
1989	172.40	0.03
1990	171.61	0.10
1991	169.74	0.03
1992	176.22	0.03
1993	175.57	0.04
1994	173.86	0.06
1995	173.35	0.06
1996	173.37	0.02
1997	169.85	0.04
1998	175.16	0.06
1999	172.89	0.10
2000	171.58	0.03
2001	170.47	0.07
2002	176.67	0.10
2003	168.23	0.08
2004	171.98	0.02
2005	168.16	0.08
2006	172.73	0.03
2007	171.36	0.06
2008	168.70	0.02
2009	177.84	0.08
2010	173.47	0.09
2011	176.40	0.04
2012	172.18	0.05
2013	170.23	0.05
2014	171.86	0.04
2015	172.81	0.03
2016	169.75	0.06

Table 2.2: The input constraints used in the SCWA for each covariate. Note that only monthly variables were available for PDO.

	D_0			D_n		
Variable	DOY	Non-Leap Year	Leap Year	DOY	Non-Leap Year	Leap Year
AIR	1	Jan. 1	Jan. 1	2	Jan. 2	Jan. 2
PDO	–	Jan.	–	–	May	–
SST	3	Jan. 3	Jan. 3	4	Jan. 4	Jan. 4
SIC	4	Jan. 4	Jan. 4	5	Jan. 5	Jan. 5

Table 2.3: Estimates and statistics of the effects of each of the four single-variable forecast models fitted with all D_{50} and environmental data through 2016.

Variable	β_0	β_1	t	R^2	F
AIR	170	-0.26	-3.41	0.25	11.62
PDO	174	-1.89	-3.63	0.28	13.17
SST	179	-1.71	-4.65	0.39	21.60
SIC	169	12.15	4.36	0.36	19.05

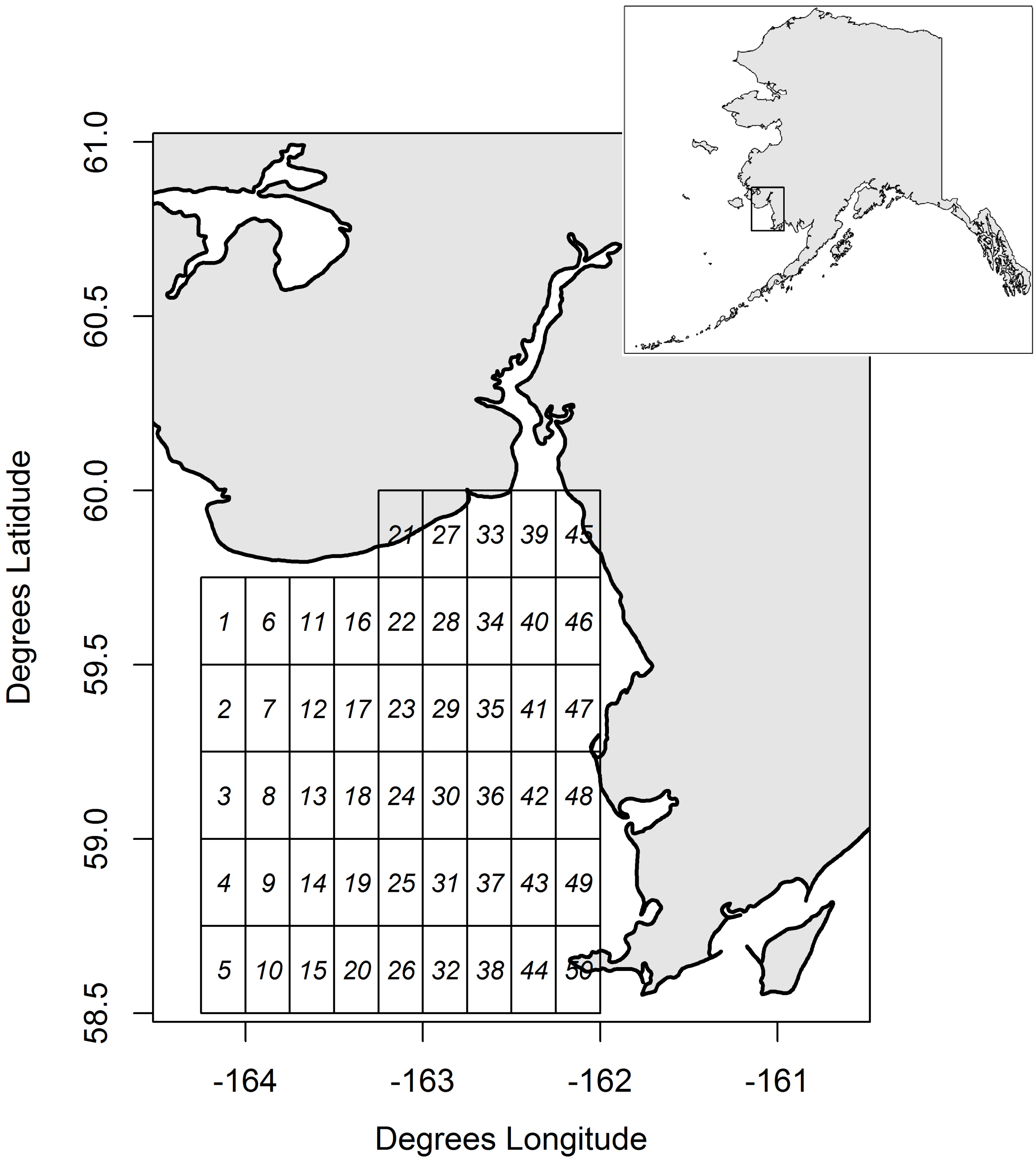

Figure 2.1: Map of Kuskokwim Bay where Chinook salmon likely stage for transition to freshwater. Shows grid cells from which daily SST values were used. Daily SIC values came from the same grid cells, though excluding grid cell 45 below due to missing values.

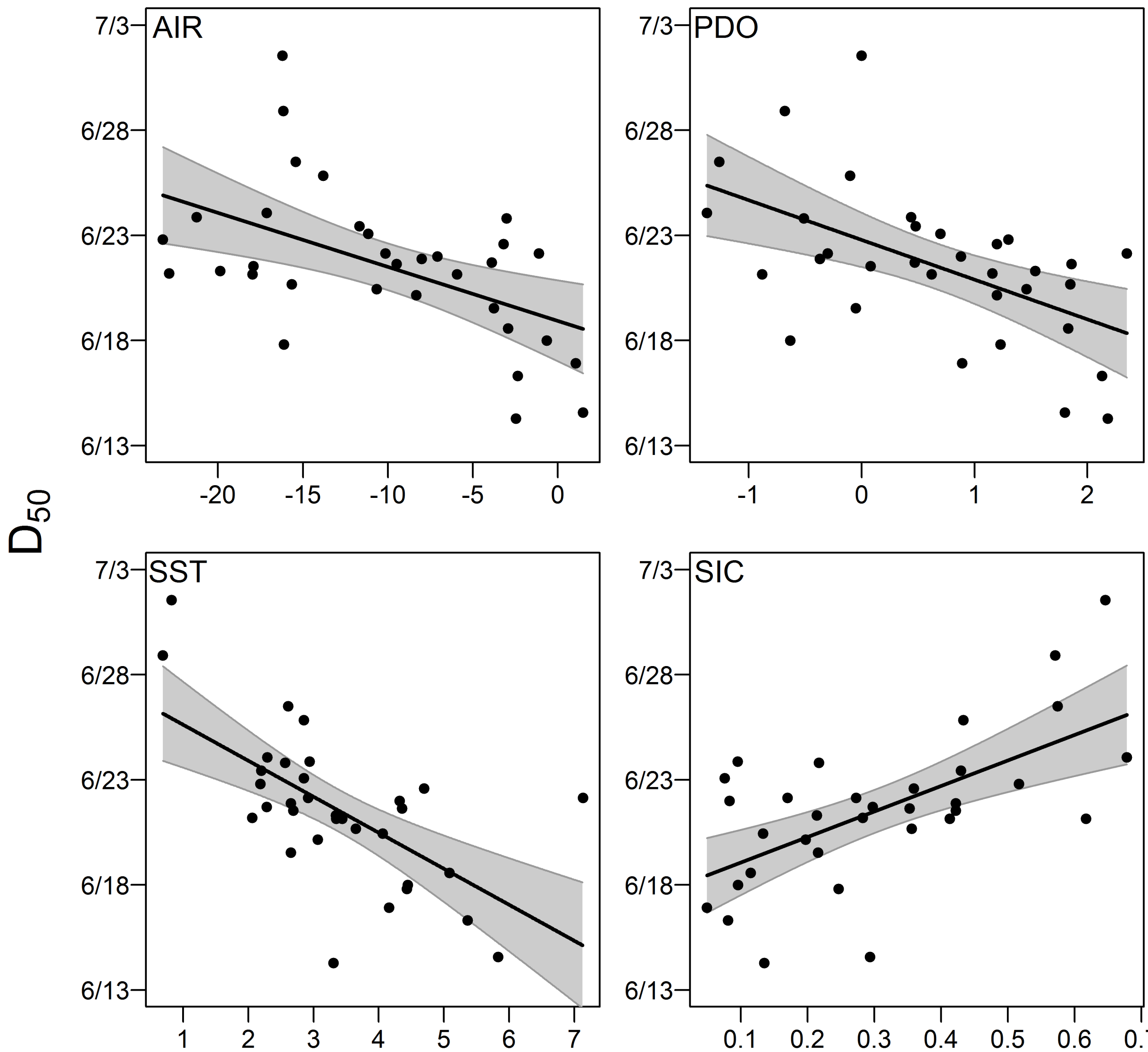

Figure 2.2: Relationships between the four single environmental variables and run timing (D_{50}) using data from optimal climate windows when 2016 was added to the training data. For illustration purposes only, gridded variables SST and SIC were combined by weighted averaging where the weight of each grid cell was assigned the $\mathrm{AIC_c}$ weight of that grid cell when grid cell-specific models were fit. Grey bands are 95% confidence intervals on the least squares line.

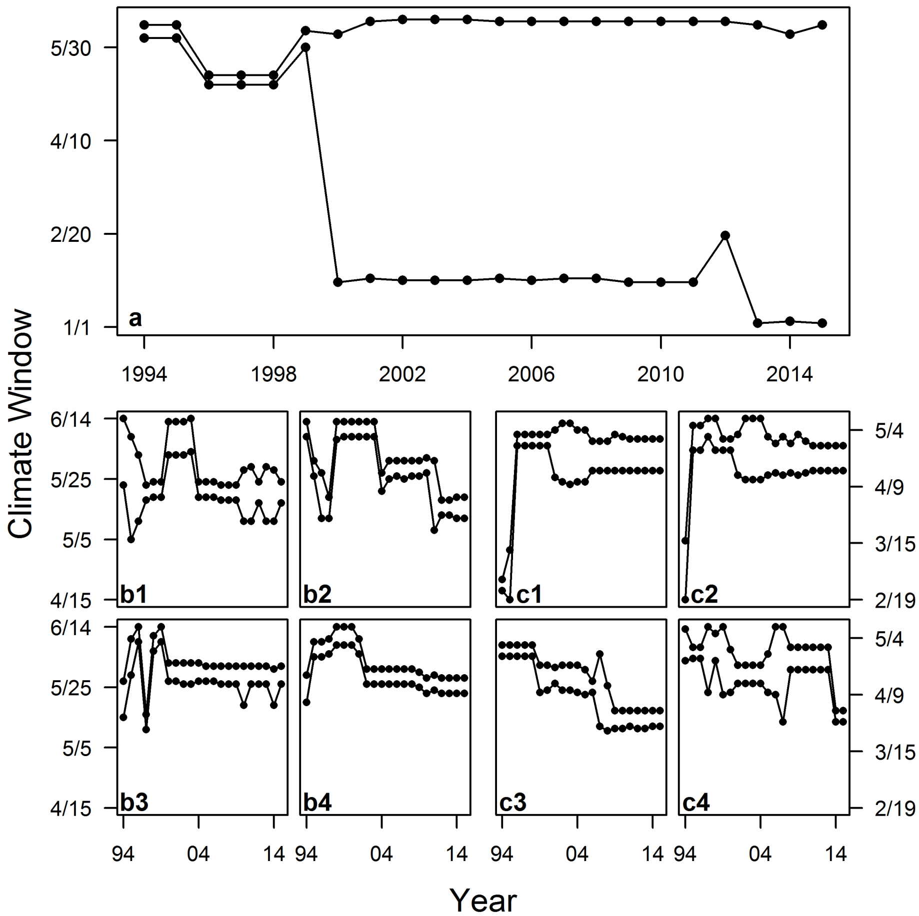

Figure 2.3: Changes in selected climate windows as training data were added in the retrospective forecasting analysis. Bottom and top lines show the first and last day of the selected climate window, respectively, as more years were added. The year axis corresponds to the selected window after including environmental and run timing data from that year in the training data. E.g., the windows shown for 2015 were used to produce the forecast for 2016. Panel (a) is Bethel air temperature, panels b1-b4 are SST windows for four sample grid cells and panels c1-c4 are SIC windows for the same four sample grid cells. Sample grid cells from Figure 2.1 shown for SST and SIC are as follows: grid cell 8 (b1, c1), grid cell 44 (b2, c2), grid cell 12 (b3, c3), and grid cell 48 (b4, c4). Selected windows for PDO are not shown because the single month of May was selected in all years.

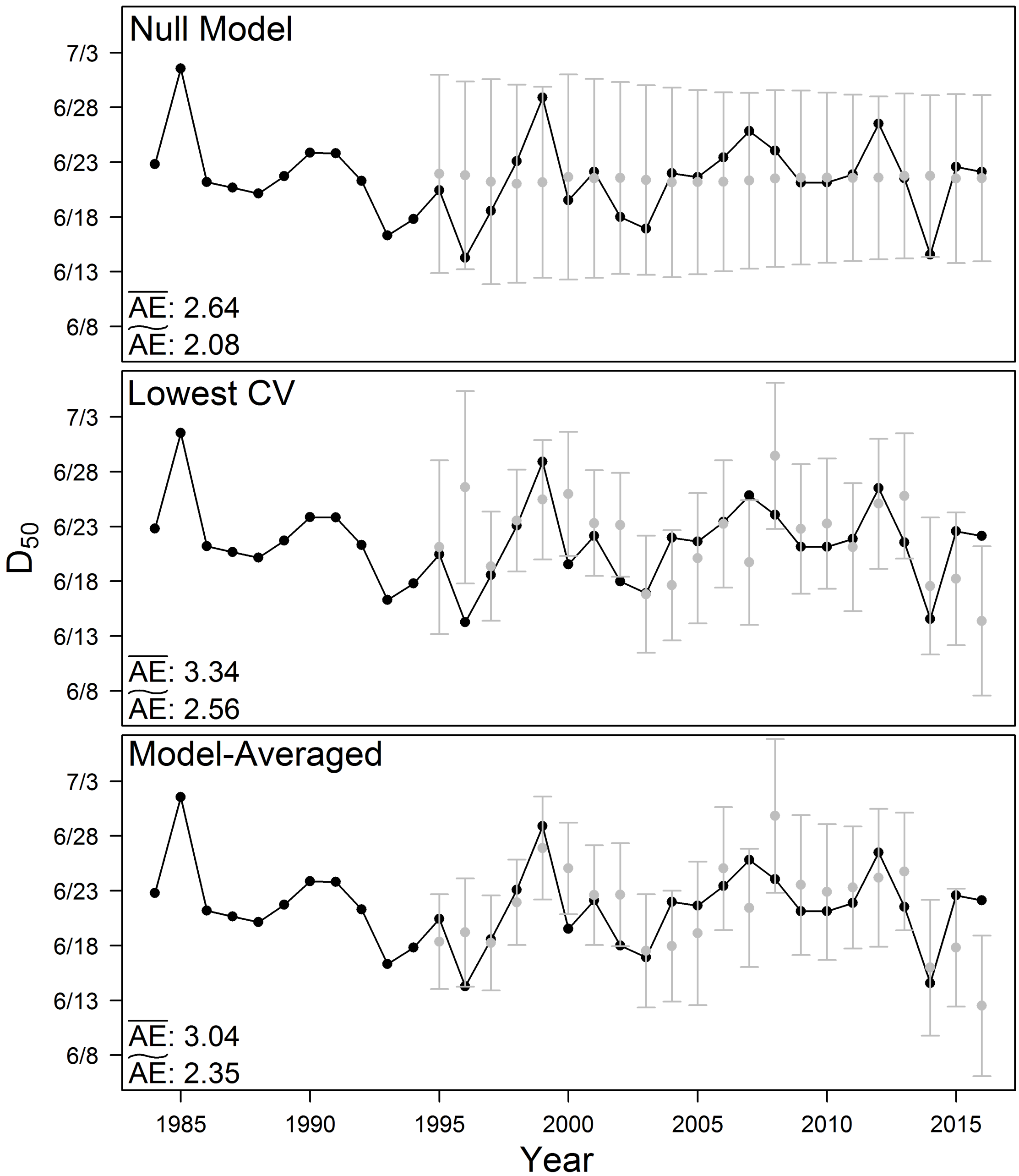

Figure 2.4: Produced forecasts under the three approaches. Black points/lines are the time series of D_{50} detected by the BTF. Grey points are out-of-sample forecasts with 95% prediction intervals shown as error bars. $\overline{\mathrm{AE}}$ and $\widetilde{\mathrm{AE}}$ are the mean and median absolute forecast errors from 1995 to 2016, respectively.

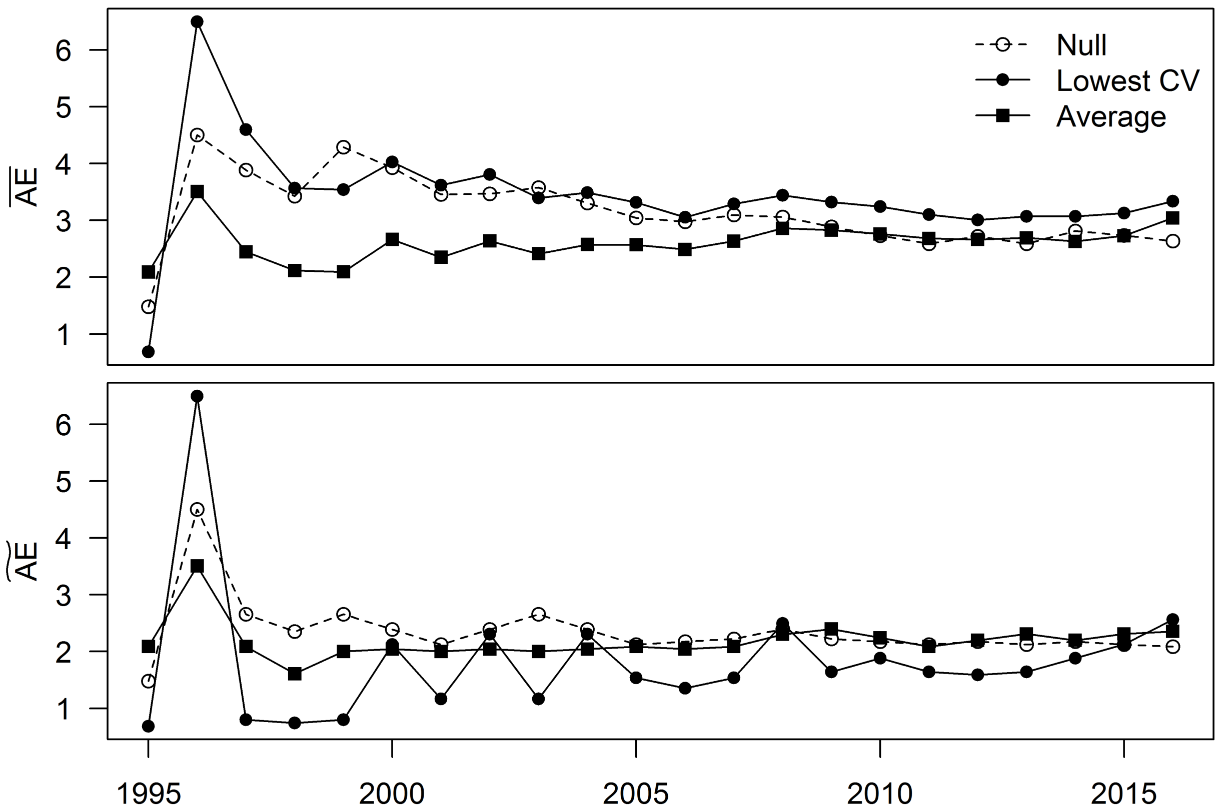

Figure 2.5: Evolution of $\overline{\mathrm{AE}}$ (mean) and $\overline{\mathrm{AE}}$ (median) absolute forecast error under the three investigated forecasting approaches. Each point is the average of absolute errors of all years before and including the corresponding year on the x-axis, starting in 1995.

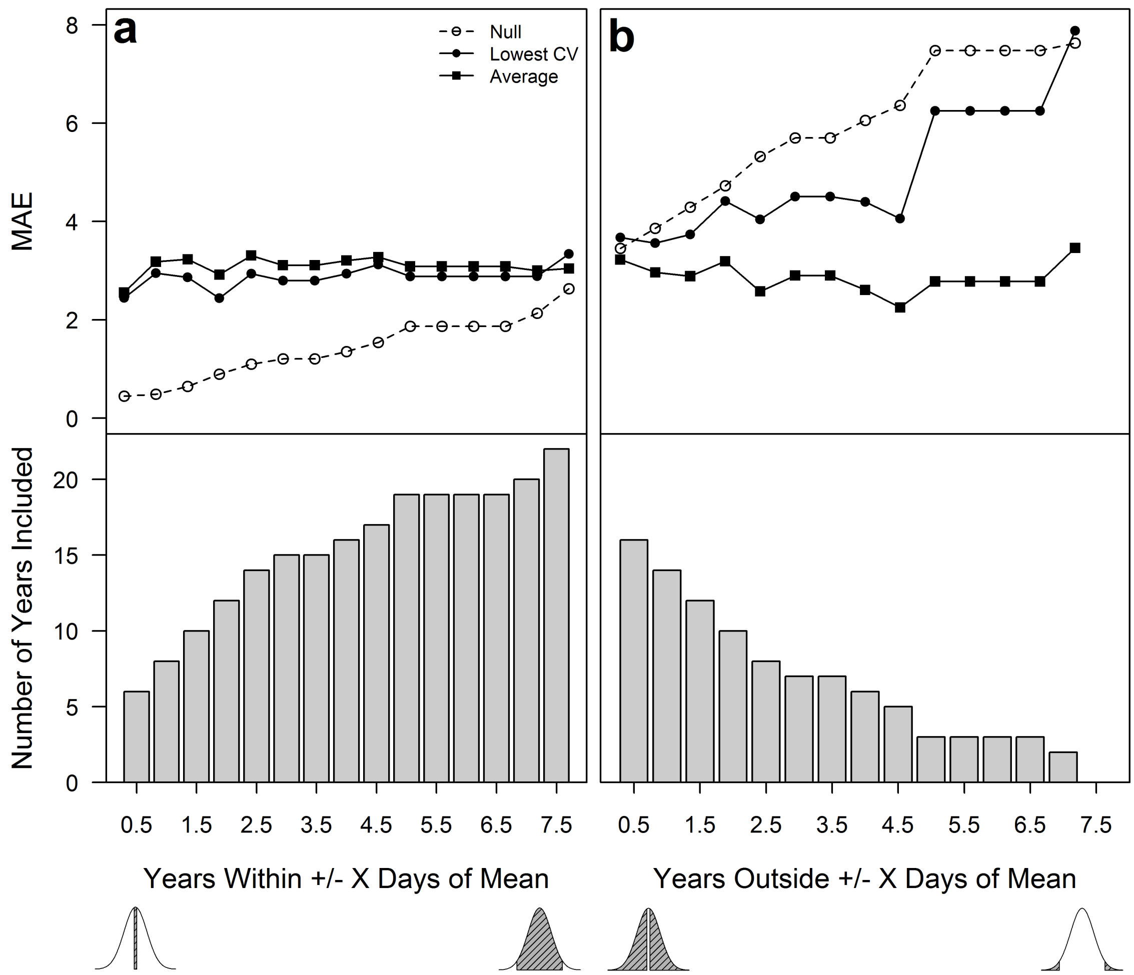

Figure 2.6: $\overline{\mathrm{AE}}$ under three forecast approaches calculated by either (a) including years with a D_{50} value within $\pm x$ days of the all-year average or (b) including years with a D_{50} value outside $\pm x$ days of average, where x is the number of days indicated on the x-axis. Bottom panels show the number of observed years in which the appropriate $\pm x$ days criterion was met. Shaded regions in the hypothetical distributions show the types of D_{50} values that were included in the calculation of $\overline{\mathrm{AE}}$. One point that may enrich inference from this figure (and is shown in the shaded normal distributions) is that panel (a) becomes more inclusive from left to right by adding years that are more dissimilar to the average in the calculation of $\overline{\mathrm{AE}}$ whereas panel (b) becomes more exclusive from left to right by removing years that are similar to the average.

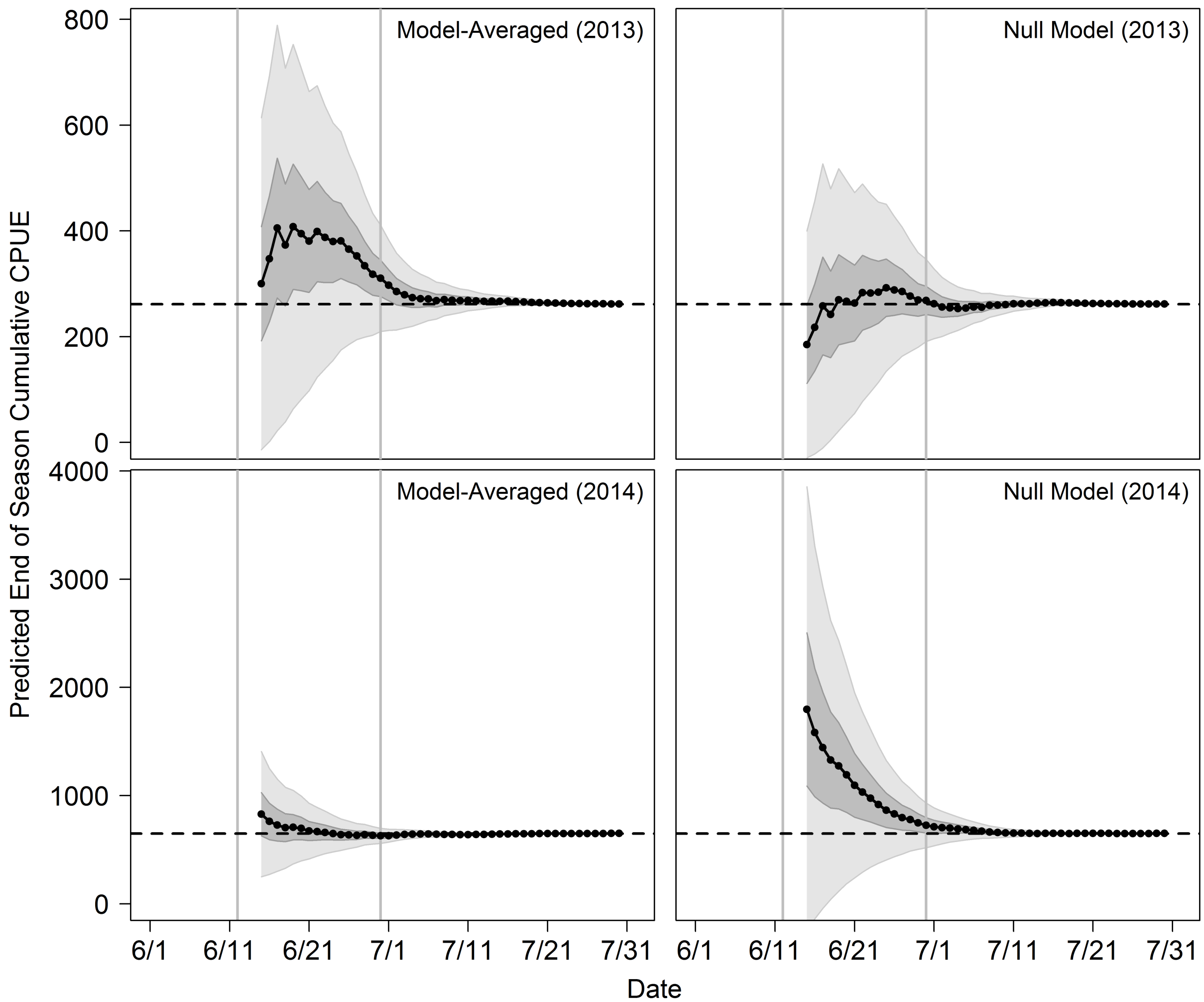

Figure 2.7: In-season predictions of end of season cumulative BTF CPUE under the model-averaged forecast using environmental variables and the forecast under the null model in 2013 and 2014. Intended to illustrate cases in which a manager would benefit from having access to the model-averaged run timing forecast model using environmental variables (2014) and when the null model would have performed better (2013). Horizontal lines are the true end of season cumulative BTF CPUE, dark grey regions are 50% confidence intervals, and light grey regions are 95% confidence intervals. Grey vertical lines indicate the period when key harvest decisions are made.

Chapter 3

Evaluation of Intra-Annual Harvest Control Rules via Closed-Loop Simulation

3.1 Introduction

Here's chapter 3. It's about in-season simulation models for management strategy evaluations.

3.2 Methods

I did some stuff.

3.3 Results

I found some stuff.

3.4 Discussion

Here's what it means.

#	Equation	Purpose/Description
1	$N_s = N_{tot}\pi_s$	Apportions total Chinook run size to subpopulations
2	$p'_{d,s} = \frac{e^{\frac{d-D_{50,s}}{h_s}}}{h_s\left(1+e^{\frac{d-D_{50,s}}{h_s}}\right)^2}$	Produces a time series of unstandardized entry timing values (logistic density function)
3	$p_{d,s} = \frac{p'_{d,s}}{\sum_d p'_{d,s}}$	Standardizes entry timing to sum to one for each Chinook subpopulation
4	$A_{d,1,s} = N_s p_{d,s}$	Populates first main stem reach with Chinook from each subpopulation
5	$A_{d,1,4} = \phi_d \sum_{s=1}^{3} A_{d,1,s}$	Populates first reach with chum/sockeye main stem abundance
6	$S_{d,r,s} = \psi_{r,s} \cdot (A_{d,r,s} - H_{d,r,s})$	Generates escapement in each reach on each day from each population
7	$A_{d+1,r+1,s} = A_{d,r,s} - H_{d,r,s} - S_{d,r,s}$	Transition main stem survivors to the next reach on the next day
8	$\text{logit}(p_{E,d,r}) = \beta_0 + \beta_1 full_{d,r} + \beta_2 stop_{d,r} + \beta_3\delta_{d-1,r,CH} + \beta_4\delta_{d-1,r,CS} + \beta + 5\phi_{d,r}$	Effort response model; $full$ and *stop* are binary indicators; δ is the fraction of needed harvest obtained for Chinook (CH) and chum/sockeye (CS), and ϕ is the local species ratio
9	$E_{d,r}p_{E,d,r}F_{d,r}$	Generates realized effort in each reach on each day
10	$H_{tot,d,r} = \min\left(1 - e^{-E_{d,r}q} \sum_{s=1}^{4} A_{d,r,s}, E_{d,r}F_{d,r}CPB_{max}\right)$	Generates total salmon harvest by reach and day

Insert Figures

Chapter 4

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

4.1 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):

$$\alpha = \frac{e^{U_{\text{MSY}}}}{1 - U_{\text{MSY}}}. \tag{4.1}$$

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 4.1, 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 4.1. 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.

4.2 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.

4.2.1 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.

4.2.1.1 Mixed-effect linear regression

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

$$R_y = \alpha S_y e^{-\beta S_y + \varepsilon_y} \tag{4.2}$$

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:

$$\log(\text{RPS}_y) = \log(\alpha) - \beta S_y + \varepsilon_y, \tag{4.3}$$

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(α)]:

$$\log(\mathrm{RPS}_{y,j}) = \log(\alpha_j) - \beta_j S_y, j + \varepsilon_y, \tag{4.4}$$

where

$$\log(\alpha_j) = \log(\alpha) + \varepsilon_{\alpha,j}, \tag{4.5}$$

and

$$\varepsilon_{\alpha,j} \sim \mathrm{N}(0, \sigma_\alpha). \tag{4.6}$$

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.

4.2.1.2 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.

4.2.1.3 Brood table reconstruction

An important point in the use of the regression-based method is in how $\text{RPS}_{y,j}$ 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 4.2) can lead to issues with this approach. Only completely observed $\text{RPS}_{y,j}$ 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 $\text{RPS}_{y,j}$, but will be provided only for years with escapement sampling for substock j. Only substocks with $\geqslant 3$ completely observed pairs of $S_{y,j}$ and $\text{RPS}_{y,j}$ were fitted.

4.2.2 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 4.1 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.

4.2.2.1 Process submodel: brood year processes

The recruitment process operates by producing a mean prediction from a deterministic Ricker (1954) relationship (Equation (4.2)) 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 (4.2), 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):

$$\bar{R}_{y,j} = \frac{\log(\alpha_j)}{\beta_j}, \tag{4.7}$$

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:

$$\bar{R}_{y,j} = \alpha_j S_{t,j} e^{-\beta_j S_{t,j}}, \tag{4.8}$$

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:

$$\log(R_{y,1:n_j}) \sim \text{MVN}\left(\log(\bar{R}_{y,1:n_j}) + \omega_{y,1:n_j}, \Sigma_R\right), \tag{4.9}$$

where

$$\omega_{y,1:n_j} = \phi\left(\log(R_{y-1,j}) - \log(\bar{R}_{y-1,j})\right), \tag{4.10}$$

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:

$$p_{y,a} \overset{\text{iid}}{\sim} \text{Dir}(\pi_{1:n_a} D). \tag{4.11}$$

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.

4.2.2.2 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:

$$N_{t,j} = \sum_{a=1}^{n_a} R_{t+n_a-a,j} p_{t+n_a-a,a}, \tag{4.12}$$

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:

$$H_{t,j} = N_{t,j} U_t v_j, \tag{4.13}$$

and escapement will be obtained as:

$$S_{t,j} = N_{t,j}(1 - U_t v_j), \tag{4.14}$$

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.

4.2.2.3 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:

$$\sigma_{\log} = \sqrt{\log(\mathrm{CV}^2 + 1)}, \tag{4.15}$$

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 ($q_{obs,t,1:n_a,j} ESS_{t,j}$).

4.2.3 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 $-0.05 \leqslant \rho < 1$ 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.

4.2.4 Analysis of Kuskokwim River substock data

4.2.4.1 Data sources

4.2.4.2 Data preparation

4.2.4.2.1 Substock escapement

4.2.4.2.2 Aggregate harvest

4.2.4.2.3 Age composition

4.2.5 Simulation-estimation analysis

4.2.5.1 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 $U_{\mathrm{MSY},j}$ and $S_{\mathrm{MSY},j}$ which (as a starting point) will be informed from random draws from the joint posterior distribution of 13 substocks from the Kuskokwim River drainage.

4.2.5.2 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.

4.2.6 Metrics of model performance

4.3 Results

I found some stuff.

4.4 Discussion

Here's what it means.

Table 4.1: 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

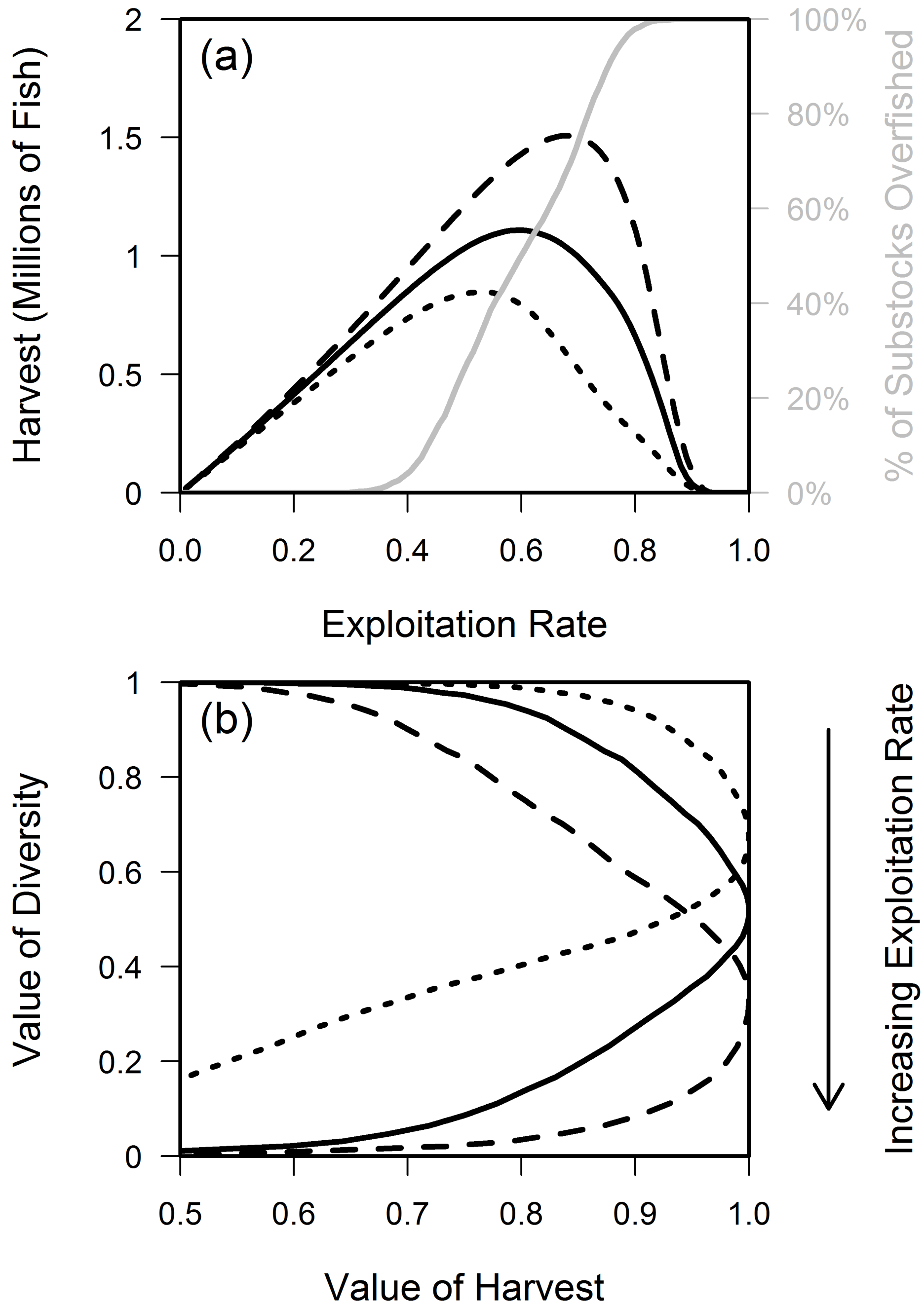

Figure 4.1: Visualization of how different types of hetergeneity in substock productivity and size influence the shape of trade-offs in mixed-stock salmon fisheries. Solid black likes are the case where stock types are split evenly among large/small and productive/unproductive stocks. Dotted black likes 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.

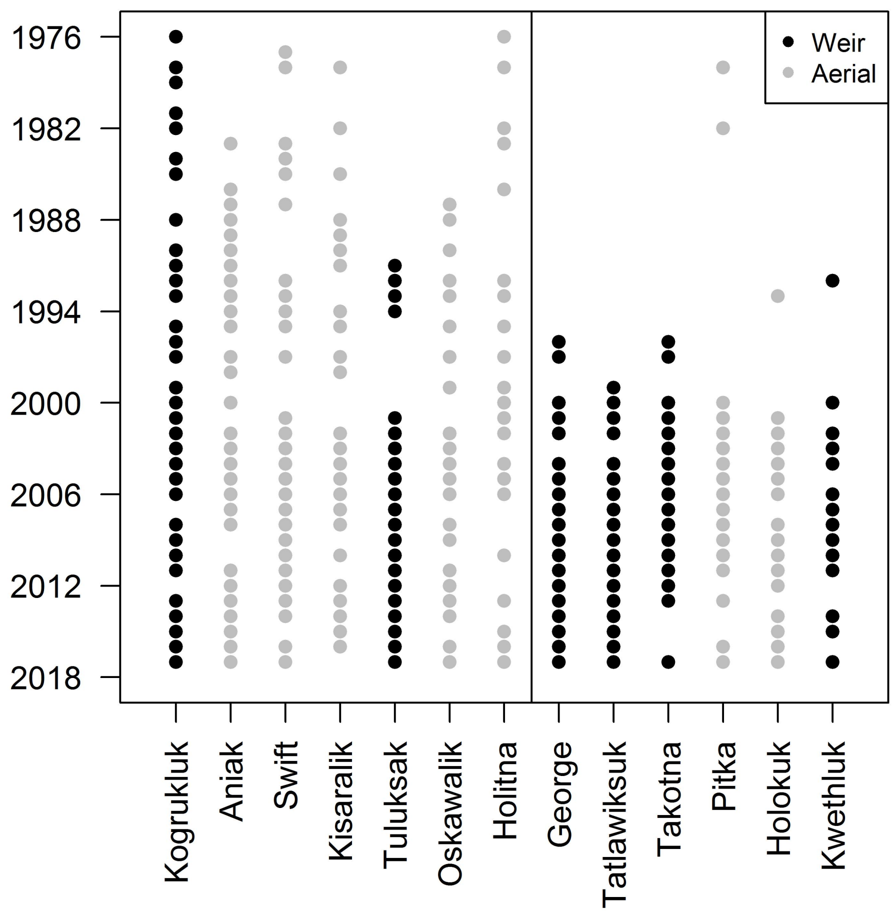

Figure 4.2: The frequency of escapement sampling for each substock sampled in the Kuskokwim River. Black points indicate years that were sampled for substocks monitored with a weir and grey points indicate years sampled for substocks monitored with aerial surveys. The vertical black line shows a break where > 50% of the years were monitored for a stock.

Chapter 5

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 2.

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.r
```

```
  return(output)
}
```

Sliding Climate-Window

Bibliography

Adkison, M. D. and Cunningham, C. J. (2015). The effects of salmon abundance and run timing on the performance of management by emergency order. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(10):1518–1526.

Anderson, J. J. and Beer, W. N. (2009). Oceanic, riverine, and genetic influences on spring chinook salmon migration timing. *Ecological Applications*, 19(8):1989–2003.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(0):40–79.

Beer, J. J. (2007). Appendix 7: Run timing of adult chinook salmon passing bonneville dam on the columbia river. In *Columbia River Salmon Passage (CRiSP) Model Monitoring and Evaluating Support Annual Report, October 1, 2006 - September 30 2007*. U.S. Department of Energy, Bonneville Power Administration, Oregon Division of Fish and Wildlife, Portland, OR.

Bolker, B. M. (2008). *Ecological Models and Data in R*. Princeton University Press. ISBN 978-0691125220.

Bromaghin, J. F. (2005). A versatile net selectivity model, with application to pacific salmon and freshwater species of the yukon river, alaska. *Fisheries Research*, 74(1-3):157–168.

Bue, B. G., Schaberg, K. L., Liller, Z. W., and Molyneaux, D. B. (2012). Estimates of the historic run and escapement for the chinook salmon stock returning to the kuskokwim river, 1976-2011. Fishery Data Series 12-49, Alaska Department of Fish and Game, Anchorage, AK.

Burger, C. V., Wilmot, R. L., and Wangaard, D. B. (1985). Comparison of spawning areas and times for two runs of chinook salmon (oncorhynchus tshawytscha) in the kenai river, alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 42(4):693–700.

Burnham, K. P. and Anderson, D. R., editors (2004). *Model Selection and Multimodel Inference*. Springer New York.

Catalano, M. J. and Jones, M. L. (2014). A simulation-based evaluation of in-season management tactics for anadromous fisheries: Accounting for risk in the yukon river fall chum salmon fishery. *North American Journal of Fisheries Management*, 34(6):1227–1241.

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.

Cooke, S. J., Hinch, S. G., Farrell, A. P., Patterson, D. A., Miller-Saunders, K., Welch, D. W., Donaldson, M. R., Hanson, K. C., Crossin, G. T., Mathes, M. T., Lotto, A. G., Hruska, K. A., Olsson, I. C., Wagner, G. N., Thomson, R., Hourston, R., English, K. K., Larsson, S., Shrimpton, J. M., and der Kraak, G. V. (2008). Developing a mechanistic understanding of fish migrations by linking telemetry with physiology, behavior, genomics and experimental biology: An interdisciplinary case study on adult fraser river sockeye salmon. *Fisheries*, 33(7):321–339.

Cooley, R. A. (1963). *Politics and Conservation: the Decline of the Alaska Salmon.* Harper and Row.

Cooperman, M. S., Hinch, S. G., Crossin, G. T., Cooke, S. J., Patterson, D. A., Olsson, I., Lotto, A. G., Welch, D. W., Shrimpton, J. M., Kraak, G. V. D., and Farrell, A. P. (2010). Effects of experimental manipulations of salinity and maturation status on the physiological condition and mortality of homing adult sockeye salmon held in a laboratory. *Physiological and Biochemical Zoology*, 83(3):459–472.

Eiler, J. H., Evans, A. N., and Schreck, C. B. (2015). Migratory patterns of wild chinook salmon oncorhynchus tshawytscha returning to a large, free-flowing river basin. *PLOS ONE*, 10(4):e0123127.

Hamazaki, T., Evenson, M. J., Fleischman, S. J., and Schaberg, K. L. (2012). Spawner-recruit analysis and escapement goal recommendation for chinook salmon in the kuskokwim river drainage. Fishery Manuscript Series 12-08, Alaska Department of Fish and Game, Anchorage, AK.

Hasler, A. D. and Scholz, A. T. (1983). *Olfactory Imprinting and Homing in Salmon.* Springer Berlin Heidelberg.

Hilborn, R. and Walters, C. J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty.* Springer US. ISBN 978-1461535980.

HODGSON, S., QUINN, T. P., HILBORN, R., FRANCIS, R. C., and ROGERS, D. E. (2006). Marine and freshwater climatic factors affecting interannual variation in the timing of return migration to fresh water of sockeye salmon (oncorhynchus nerka). *Fisheries Oceanography*, 15(1):1–24.

Keefer, M. L., Peery, C. A., and Caudill, C. C. (2008). Migration timing of columbia river spring chinook salmon: Effects of temperature, river discharge, and ocean environment. *Transactions of the American Fisheries Society*, 137(4):1120–1133.

Keefer, M. L., Peery, C. A., Jepson, M. A., and Stuehrenberg, L. C. (2004). Upstream migration rates of radio-tagged adult chinook salmon in riverine habitats of the columbia river basin. *Journal of Fish Biology*, 65(4):1126–1141.

Linderman, J. C. and Bergstrom, D. J. (2009). Kuskokwim management area: Salmon escapement, harvest, and management. 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 541–599, Bethesda, MD.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C. (2017). A pacific interdecadal oscillation with impacts on salmon production. *Bulletin of the American Meterological Society*, 78(6):1069–1079.

Mundy, P. R. and Evenson, D. F. (2011). Environmental controls of phenology of high-latitude chinook salmon populations of the yukon river, north america, with application to fishery management. *ICES Journal of Marine Science*, 68(6):1155–1164.

O'Malley, K. G., Ford, M. J., and Hard, J. J. (2010). Clock polymorphism in pacific salmon: evidence for variable selection along a latitudinal gradient. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701):3703–3714.

Quinn, T. P., Unwin, M. J., and Kinnison, M. T. (2000). EVOLUTION OF TEMPORAL ISOLATION IN THE WILD: GENETIC DIVERGENCE IN TIMING OF MIGRATION AND BREEDING BY INTRODUCED CHINOOK SALMON POPULATIONS. *Evolution*, 54(4):1372–1385.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, 20(22):5473–5496.

Salinger, D. H. and Anderson, J. J. (2006). Effects of water temperature and flow on adult salmon migration swim speed and delay. *Transactions of the American Fisheries Society*, 135(1):188–199.

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.

van de Pol, M., Bailey, L. D., McLean, N., Rijsdijk, L., Lawson, C. R., and Brouwer, L. (2016). Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution*, 7(10):1246–1257.

Walters, C. J. (1986). *Adaptive Management of Renewable Resources*. Blackburn Press. ISBN 978-1930665439.

Walters, C. J. and Holling, C. S. (1990). Large-scale management experiments and learning by doing. *Ecology*, 71(6):2060–2068.

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

Wolfe, R. J. and Spaeder, J. (2009). People and salmon of the yukon and kuskokwim drainages and norton sound in alaska: Fishery harvests, culture change, and local knowledge system. 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 349–379, Bethesda, MD.

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