Modeling the Population Dynamics of Herring in the Prince William Sound, Alaska

Melissa L. Muradian

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington
2015

Committee:
Trevor A. Branch
Melissa A. Haltuch
André E. Punt
Ian J. Stewart

Program Authorized to Offer Degree:
Quantitative Ecology and Resource Management

University of Washington

Abstract

Modeling the Population Dynamics of Herring in the Prince William Sound, Alaska

Melissa L. Muradian

Chair of the Supervisory Committee:
Associate Professor Trevor A. Branch
School of Aquatic and Fishery Sciences

The Pacific herring (*Clupea pallasii*) population in Prince William Sound, Alaska crashed in 1992-93 and has yet to recover, affecting food web dynamics in the Sound and impacting Alaskan communities. To help researchers design and implement the most effective monitoring, management, and recovery programs, a Bayesian assessment of Prince William Sound herring was developed and conducted by reformulating the current model used by Alaska Department of Fish and Game to manage this population. The pre-season spawning biomass of herring age-3 and older in 2013 was estimated to be between 12,150 and 31,740 mt (95% credibility interval) with a median of 19,410 mt. There was a 54% probability that biomass in 2013 was below the management limit used to regulate fisheries in Prince William Sound. This work could be the basis for managing herring in Prince William Sound in combination with a decision rule that explicitly considers uncertainty.

A Monte-Carlo simulation study, using the Bayesian Prince William Sound herring assessment model, was conducted to determine which historical sampling programs have provided the most valuable data for forecasting biomass. Tradeoffs were explored between the cost of running each survey and the improvement in model performance due to the inclusion of that survey's data. The disease survey (which is relatively cheap and collects an index of additional mortality due to disease) and the diver survey (which is relatively expensive and collects an absolute index of abundance) were found to be the most valuable sampling programs. For $10,000 a year the disease survey reduces bias and imprecision in the forecast by 34% on average, increases model reliability by 22%, and decreases by 31% the probability of a false management conclusion when regulating the fishery. For $350,000 a year the diver survey reduces bias and imprecision in the forecast by 12% on average, increases model reliability by 6%, and decreases the probability of a false management conclusion by 23%. This work is directly useful to Pacific herring assessment scientists and provides critical information about how to prioritize research and monitoring efforts to better understand the past and future ecology of Prince William Sound herring.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I extend my heartfelt thanks to my advisor and chair, Trevor A. Branch, for sharing with me his support, guidance, and inspiration; I cannot imagine a better research advisor. Special thanks to Andre E. Punt whose energy, interest, and direction have been incredibly helpful and seemingly limitless. I cannot overstate the profound impact Trevor's and Andre's outlook and expertise have had on these, my formative years as a scientist. I would also like to thank my committee members Ian J. Stewart and Melissa A. Haltuch for their guidance and criticism, which have helped shape and improve my research.

Special thanks to Steve Moffitt from the Alaska Department of Fish and Game for tirelessly answering my numerous model- and data-related questions and for offering insight and background about this extraordinary population. Thanks to Rich Brenner from the Alaska Department of Fish and Game and to Michele Buckhorn from the Prince William Sound Science Center for sharing their knowledge with me and for providing me an invaluable opportunity to first, witness and better understand several of the survey collection programs and second, to actually hold a herring! Thanks to W. Scott Pegau from the Prince William Sound Science Center and to Paul Hershberger from the U.S. Geological Survey for providing insight as well as survey cost information, and thanks to Sherri Dressel from the Alaska Department of Fish and Game and Steve Martell from the Pacific Halibut Commission for brief yet important discussions about the analysis. This work was greatly aided by the preceding research and model development done by Peter F.J. Hulson, to whom I extend my thanks and appreciation.

The Exxon Valdez Oil Spill Trustee Council, grant 13120111-Q, funded this work and I am thankful to the Council for their support.

INTRODUCTION

Stock assessment is the practice of fitting a population dynamics model to data, using an assumed statistical relationship between predicted and observed values, to estimate stock status and regulate harvesting. In practice, an assessment model is used to estimate life-history parameters to set catch limits or harvest rates based on the size and productivity of the population. Therefore, accurate assessment model estimates are necessary to prevent overfishing or underutilization of a managed stock.

There are several components that contribute to a useful stock assessment model and researchers continue to test, revise, and improve current understanding and intuition regarding complex models. The three key components of an assessment model are the data, the assumed population dynamics, and the assumed statistical relationship between the two. In this thesis I increase the utility of the stock assessment model used by the Alaska Department of Fish and Game (ADF&G) to manage Pacific herring (*Clupea pallasii*) in Prince William Sound, Alaska first by updating the statistical basis for parameter estimation and inference in the assessment from weighted least squares to a Bayesian framework, and then by conducting a simulation study to determine which data types are the most valuable for estimating accurate, precise, and reliable population quantities, such as biomass, mortality, and recruitment.

Pacific herring are the principal pelagic forage fish in Prince William Sound, Alaska. They are valuable prey for marine birds, marine mammals such as sea lions and humpback whales, and piscivorous fish such as pink salmon and Pacific cod. Herring have been commercially harvested from the Sound for human use and consumption for over a century, and although a record is lacking, native harvests probably precede these. Thus, herring are central nodes in the Prince William Sound ecosystem, and were historically an important part of the economy in the Sound.

In 1989, the Prince William Sound was the site of the Exxon-Valdez oil spill. The spill occurred at the end of March—the start of the herring-spawn season—when hundreds of millions of herring migrate from the open ocean into the Sound to spawn. This enormous influx of herring into the Sound during spawning attracts numerous predators, so that permanent and transitory species experienced negative effects from the spill (Piatt et al., 1990; Peterson, 2001; Monson et al., 2000). Nonetheless, herring biomass remained high for several years before declining in 1992, and by 1993 the population had dropped to such low abundance that all commercial herring fisheries in Prince William Sound were closed. The fishery was briefly re-opened following a period of perceived recovery, but the herring population again collapsed only a few years later in 1999 and has not yet recovered. Today, the herring population remains at low abundance and the herring fishery has been closed for nineteen consecutive years.

Since 1993, herring have been intensively monitored as part of ongoing research programs aimed at recovery of the numerous species and habitats in the Sound impacted by the Exxon-Valdez oil spill. Due to their pivotal role in the food web dynamics of the ecosystem, herring recovery is expected to assist in the rebuilding of many other non-recovering species impacted by the spill. A key component of these monitoring and recovery efforts is the stock assessment model used by ADF&G to manage, and assess the status of, Prince William Sound herring. The assessment model integrates multiple types of observations to predict the size of the Prince William Sound herring population each year. As part of the herring research and monitoring efforts, the following chapters outline my work to update the assessment model from a weighted least squares to a Bayesian framework and use it to determine the relative value of each sampling program for estimating accurate, precise, and reliable population quantities.

Chapter 1. A BAYESIAN STOCK ASSESSMENT OF PACIFIC HERRING IN PRINCE WILLIAM SOUND, ALASKA

1.1 INTRODUCTION

1.1.1 *General background*

Lower-trophic level species form an integral part of marine food web dynamics by transferring production from plankton to larger predatory fish, marine mammals, and sea birds (Smith et al., 2011; Cury et al., 2011). Globally, low-trophic level species, such as sardine, herring, and sand lance, account for more than 30% of fisheries production (Smith et al., 2011). Therefore maintaining the abundance of small pelagic "forage" fish is inextricably tied to both ecological and anthropocentric food security.

Pacific herring (*Clupea pallasii*) are the principal pelagic forage fish in Prince William Sound (PWS), Alaska. In the Sound, the herring population provides valuable prey for marine birds, marine mammals such as sea lions and humpback whales, and piscivorous fish such as pink salmon and Pacific cod (Thomas and Thorne, 2001; Thomas and Thorne, 2003). Pacific herring have been harvested by humans for thousands of years and PWS herring have been commercially harvested for over a century (Funk and Sandonne, 1990; Pete, 1990); this population has been at such high levels of abundance in the past as to sustain a catch of over 40,000 t for five consecutive years in the 1930s (Fig. 1.1). Therefore, herring in Prince William Sound are key components of the surrounding ecosystem and were historically an important part of the local economy.

In 1989, the PWS was the site of the Exxon-Valdez oil spill, which occurred when the eponymous oil tanker ran aground off Bligh Island, spilling millions of gallons of crude oil into the Sound, and soiling over 1,000 miles of Alaskan coastline. The spill occurred at the end of March, the start of the herring-spawn season when hundreds of millions of herring migrate from more open areas to shallower, coastal waters of Prince William Sound to spawn (Haegele and Schweigert, 1985), thus herring were adversely affected by the spill that and the following year (Brown and Baker, 1998; Hose et al., 1996; Marty et al., 1999; Norcross et al., 1996). During their pre-spawn activity, herring form aggregations of millions of fish, which makes them easy targets for predators within the Sound and attracts predators from outside of the Sound (Thomas and Thorne, 2003). For example, humpback whales migrate into the Sound to prey on the pre-spawn herring aggregations (Teerlink et al., 2014). Therefore, the herring-spawn coincides with increased species diversity and activity within PWS, and as a result, many permanent and transitory species experienced negative effects from the Exxon-Valdez oil spill (Piatt et al., 1990; Peterson, 2001; Monson et al., 2000).

Even though herring experienced negative effects from the spill, total biomass remained high until the population collapsed in 1992–3 (Quinn et al., 2001). Since the spill occurred right before the fishery season began in 1989, no spring removals were taken that year, but fishing resumed in fall 1989. Over 10,000 t of total catch was taken in 1990, and catches in 1991 and 1992 were higher than any during any of the previous thirty-five years of fishing (Fig. 1.1). Herring were expected to have record high recruits in 1993, but harvest was limited in 1993 due to extremely low observed spawning biomass (Hulson et al., 2008; Quinn et al., 2001; Thorne and Thomas, 2008). In retrospect, this year marked the historic collapse of the herring population that underwent years of causal investigation. The herring fishery was closed in 1994, 1995, and the spring of 1996 due to low biomass, opened back up in the fall of 1996–1998 due to trends of increasing biomass, and closed once more in 1999 due to insufficient biomass (Figs 1.1, 1.2). Since 1999, the Sound has been closed to herring fishing due to continued low abundance, which has cost Alaskan communities approximately $230 million dollars in lost income (Kopchak, 2013).

Once it was clear that herring were not recovering from the collapse in 1993, research programs were initiated to intensively monitor the post-spill abundance and fitness of herring, as well as collect data on physical and environmental factors to guide restoration efforts. While recovery of this population is important in its own right, rebuilding it has become a focus of the overall restoration efforts in the Sound due to herring's pivotal role in that ecosystem's food web dynamics; the hope is that a herring recovery will assist in rebuilding many other non-recovering species affected by the Exxon-Valdez oil spill.

1.1.2 *Managing the herring fisheries in the Prince William Sound*

The Alaska Department of Fish and Game (ADF&G) has managed PWS herring since 1973. They collect data on herring and use an assessment model to forecast the following year's spawning biomass. This forecast quantity, called pre-fishery run biomass, is the projected biomass at the start of the spring fishing season, and is used to set harvest rates using a lower and an upper regulatory biomass threshold. If pre-fishery run biomass falls below the lower regulatory threshold of 22,000 short tons (19,958 mt), then the fishery is closed that year. The lower regulatory threshold is based on a minimum spawning biomass threshold of 25% of the potential spawning biomass from an un-fished state (Botz et al., 2010). If forecasted biomass is between the lower regulatory threshold and an upper regulatory threshold of 42,500 short tons (38,555 mt), then a harvest rate increasing linearly from 0.0 yr^{-1} to 0.2 yr^{-1} is set; and if forecasted biomass is above the upper regulatory threshold, then the harvest rate is set to 0.2 yr^{-1}. The resulting catch limit is then divided among the four major herring fisheries: the gillnet fishery, the pound spawn-on-kelp fishery, the sac-row purse-seine fishery, and the food/bait fishery.

The food/bait fishery was conducted during the fall, and almost exclusively used purse-seine gear (and rarely trawl gear) to harvest whole herring for use as bait on hook and line gear or for human consumption. The remaining three fisheries ran during the spawning season in the spring. The purse-seine sac-roe fishery harvested herring egg or roe sacs using purse seine and gillnets and was actively managed to obtain the highest valued product possible through monitoring of ripening females and timing of the main spawning event. During spawning, females deposit their eggs on marine vegetation in the inter-tidal zone, and these egg-encrusted kelp fronds are harvested by spawn-on-kelp fisheries (Baker and Brady, 1992; Haegele and Schweigert, 1985). The PWS pound fishery is a type of spawn-on-kelp fishery that involves impounding mature herring in a net with suspended kelp, until they spawned. Herring that survive being impounded are released and the kelp with eggs attached, called komochi konbu, is harvested and sold.

1.1.3 *Past development of the assessment model and a brief description of key assumptions and data*

Beginning in 1973, the agency initially estimated the forecast biomass using an index of male spawning biomass (called mile-days of milt) collected by aerial surveys until 1988, when an age-structured assessment model (known as the ASA Model) was developed. This used the same structure as the assessment model used today, which has been expanded and updated to incorporate more data types and more realistic assumptions. The initial version of the ASA Model used catch-at-age data, along with the aerial survey information, but also incorporated annual numbers of eggs deposited (called egg deposition data) as an absolute index of female spawning biomass (Funk and Sandonne, 1990). Later, an assumption of increased levels of mortality from high levels of disease starting in 1992 was incorporated into the ASA model to fit observed trends in the population (Marty et al., 2003; Quinn et al., 2001). An index of hydroacoustic biomass was formally incorporated into the ASA model to address conflicting trends between the mile-days of milt and egg-deposition data over 1987–91 (Hulson et al., 2008) (Figs 1.3). Prior to this, observed hydroacoustic biomass data were not used for model fitting, but were instead used as a lower constraint on estimated pre-fishery run biomass. Even though evidence is lacking for a Ricker stock-recruit relationship into PWS herring, this relationship was recently introduced to the ASA model to stabilize estimates of recruitment (Hulson et al., 2008). The assumed Ricker relationship operates as a penalty to keep estimated age-3 abundance close to the assumed recruitment relationship and to avoid recruitment from converging towards zero in some years. Details of the PWS herring ASA Model are included in the Materials and Methods section.

1.1.4 *Challenges to modeling the population dynamics of Prince William Sound herring*

A number of competing hypotheses have been proposed to explain the recent population trends of PWS herring. Proposed explanations related to direct and indirect effects from the oil spill include: toxicity related to ingestion of hydrocarbons, recruitment failure related to increased egg loss, and poor over-wintering condition related to poor growth and low plankton levels (Hulson et al., 2008; Pearson et al., 2012). Some debate exists as to whether these factors related to the oil spill caused the herring collapse of 1992–93, and even more speculation surrounds the causes of the continued low abundance. Potential explanations include persistent effects of the oil spill, increased predation coinciding with increasing populations of humpback whales, increased disease mortality, recruitment failure, and increased interspecific competition with millions of introduced pink salmon (Pearson et al., 2012). In relation to the latter potential factor, hatchery pink salmon numbers more than doubled to over 500 million fry released upriver of the Sound in the same year as the oil spill (Fig. 1.2). This is important because juvenile pink salmon compete with herring for food, and juvenile pink salmon eat age-0 and age-1 herring. The former interaction is of particular importance since Pacific herring show evidence of density-dependent growth (Reum et al., 2013), meaning limited food resources can reduce fat stores, leading to increased over-winter mortality.

What caused the crash? Was it toxicity or habitat loss from the oil spill or an increase in disease mortality? Are these the same factors contributing to recent low abundance or have other pressures, such as interspecific competition or increased predation, been introduced? In order to explore these questions, assessment scientists need the best available tools to reliably estimate population metrics, such as spawning biomass, and to test hypotheses about what factors are influencing the continued low abundance. Therefore, this study implemented several updates to the ASA Model including recasting it in a Bayesian framework, testing assumptions, and investigating evidence of a retrospective pattern.

The ASA model currently employs a weighted least squares estimation framework that provides point estimates of annual pre-fishery run biomass (Fig. 1.2), meaning managers make decisions based on the biomass forecast without an indication of the variance associated with that model estimate. Furthermore, the weighting on each term in the sum of squares does not utilize estimates of sampling error (when provided), but is chosen to reflect consensus views about the relative precision of each dataset (Hulson et al., 2008). Here we expand the model into a maximum likelihood (ML) and Bayesian framework. These offer methods to statistically estimate the uncertainty involved in estimates of management quantities, such as projected spawning biomass, by replacing these subjective likelihood weights with theoretically valid estimates of uncertainty based on the assumed form of the distribution.

Bayesian methods, in particular, offer several advantages over weighted least squares including easily interpretable estimates of variance in the form of probability distributions (posteriors), which enable managers to report probabilities associated with abundance estimates and alternative management scenarios useful in decision analysis and management strategy evaluation (McAllister et al., 1994; Punt and Hilborn, 1997; Smith et al., 1999). In addition to using information contained in the data, Bayesian theory formally allows incorporation of expert opinion, or information from other stocks or species, into the model fitting process through the use of prior probability distributions. Moreover, Bayesian decision analysis can incorporate the full range of uncertainty across models and parameters. This is in contrast to conditional ML methods that may fail to incorporate the actual variability of several parameters (Francis, 1992;

Punt and Hilborn, 1997). Therefore, this study implements the PWS herring ASA model in a Bayesian framework to provide the following benefits: 1) a conceptually simple definition of uncertainty via credible intervals, and 2) the assignment of probabilities to hypotheses concerning population dynamics and state of the fishery.

1.2 MATERIALS AND METHODS

Fifteen data types were used in this study (Table 1.1), including weight-at-age, age compositions, a milt index, a hydroacoustic biomass index from the ADF&G surveys, and a biomass index from the Prince William Sound Science Center (PWSSC) hydroacoustic survey. The longest time series spans 1980–2012, but many series have years or sets of years of missing data (Table 1.1). I developed the Bayesian form of the ASA model in AD Model Builder (Fournier et al., 2012). Using a single MCMC chain, posterior distributions were obtained for all parameters and key model outputs, such as pre-fishery run biomass and fishing mortality. The following sections describe the data, the population dynamics and modeling assumptions, the forms of the likelihood functions and prior distributions, and convergence criteria used when fitting the Bayesian ASA model.

1.2.1 *Fishery-dependent data*

Catch data over the period 1980–1999 from all four major herring-fishing fleets in the Sound were used (Fig. 1.4). Catch-at-age data (in millions of fish) came from the gillnet fishery (Table index 1.1.1), the pound-utilization spawn-on-kelp fishery (Table index 1.1.2), and the food/bait fishery (Table index 1.1.3). The food/bait fishery was conducted during the fall, and the former two fisheries ran during the spawning season in the spring. We assume that 75% of impounded fish die during the spawn-on-kelp fishery, and consequently set $\rho_k = 0.75$ (Table index 1.4.2) (Hulson et al., 2008). Annual data from the purse seine fishery is in the form of total yield, in metric tons, and catch at-age proportions (Table index 1.1.8 and 1.1.11).

1.2.2 *Fishery-independent data*

Fishery-independent surveys include a hydroacoustic survey conducted by PWSSC and several surveys conducted by ADF&G during the herring-spawn (described in the following paragraphs). ADF&G uses purse-seine gear to sample spawning assemblages and collects information on age compositions (Table index 1.1.9), weight-at-age (Table index 1.1.7), and the proportion of spawning herring that are female (Table index 1.1.10).

Two separate indices of acoustic abundance are used in the model; one collected by ADF&G and one collected by the PWSSC (Table indices 1.1.14–15). The PWSSC began collecting herring acoustic biomass data in 1993, and the first two years of data in this series were collected in fall. The remaining data were sampled during the herring spawn in the spring. The two series have been combined into a single time series, since PWSSC conducted seasonal acoustic surveys during 1995–7 and found no significant difference between the fall and spring results (Thorne and Thomas, 2008).

This study includes only hydroacoustic ADF&G observations that are independent from the PWSSC survey (Table index 1.1.14). Even though the ADF&G began conducting acoustic surveys every other year starting in 1997, data used by the ADF&G were a combination of their observations and those of the PWSSC until 2004 (Hulson et al., 2008; Thomas and Thorne, 2003). Independent ADF&G hydroacoustic biomass estimates are available for years 2005–2009.

Diver surveys were conducted by the ADF&G in 1984 and 1988-1992 to collect data used to measure herring fecundity-at-age (Table index 1.1.6) and in 1984, 1988-1992, and 1994-1997 to estimate the total numbers of eggs deposited in the inter-tidal spawning beds (the egg deposition index) (Table index 1.1.12). We converted the 95% confidence intervals from the egg deposition data into coefficients of variation (Table index 1.1.13). There is high uncertainty associated with this index due to inter-annual changes in spawning location within the Sound and since egg loss continuously occurs due to wave-action and predation, while diver surveys of spawning beds can take place anywhere from days to weeks after spawning events (Rooper et al., 1999). In the model, an important assumption is that egg deposition is directly related to absolute spawning biomass after accounting for fecundity-at-age data and numbers-at-age estimates (Table index 1.4.21).

The ADF&G aerial milt survey from 1980–2012 is the final dataset used (Table index 1.1.17). During spawning events, aerial surveys fly along the spawning sites and measure the linear extent of milt clouds in miles of corresponding coastline per day, hence the mile-days units. This metric was developed in 1987 to address the issue of residence uncertainty in the aerial biomass estimates historically used by the ADF&G to estimate spawning biomass (Funk and Sandonne, 1990). Residence uncertainty refers to issues inherent in measuring peak spawning biomass using aerial surveys that require assumptions about the timing of fish movement in and out of the spawning beds during the weeks-long spawning period. Mile-days of milt are a key index of total herring biomass that closely track the hydroacoustic biomass estimates (Fig. 1.3).

When the four indices of biomass (milt, egg deposition, and the two acoustic indices) are overlaid, complementary and conflicting trends can easily be identified (Fig. 1.3). All indices display reasonably common trends after 1993. The indices of milt and egg deposition start out at relatively high levels during the late 1980s and drop to lower levels during the mid-1990s, but they show opposing trends during the years 1988–1992. Milt shows a tremendous increase that peaked in 1989 and then sharply decreased over the next six years. This is in contrast to the trend in the egg deposition data, where moderate numbers of eggs were deposited in 1988–9, but numbers sharply increased over the next three years. This is a well-documented conflict in the PWS herring data (Hulson et al., 2008).

1.2.3 *Disease survey data*

Infection data are included in the model for the protozoan parasite *Ichthyophonus hoferi* and the North American strain of viral haemorrhagic septicemia virus (VHSV). High levels of infection from *I. hoferi* in 1989-1993 and VHSV in 1993 were found in PWS herring (Kocan et al., 1996; Hulson et al., 2008; Marty et al., 1999; Marty et al., 2010), and these diseases were found to significantly affect herring mortality at the individual level (Marty et al., 2003). This information combined with the 1992/93–population crash influenced the initialization of a systematic survey in 1994 to collect data on diseases affecting PWS herring. The herring assessment model integrates information from these diseases as two series: the first is an index for *I. hoferi* infection (Table indices 1.1.4 and 1.1.5) and the second is a index that combines direct evidence of VHSV and evidence of ulcers related to filamentous bacteria. Previous studies concluded that a combined VHSV/ulcer index better explains survival trends and fluctuations in associated annual mortality since ulcers can be a surrogate indicator of VHSV infection (Marty et al., 2010).

1.2.4 *Model formulation: population dynamics, and prior distributions*

The PWS herring population is managed as a fully mixed population under the assumption of no immigration or emigration from populations outside PWS, since geographic barriers isolate PWS herring from other spawning populations (Funk and Sandonne, 1990). Biomass losses occur due to either fishing events within the Sound (there is no evidence of fishing on this population outside of the Sound) or natural mortality (including disease), while increases occur through recruitment and growth in body weight with increasing age.

In common with the current ADF&G assessment model, the Bayesian model is a statistical catch-at-age model (Quinn and Deriso, 1999). Three-year-old herring are the first age class that is observed to be present in large numbers on the spawning grounds and Pacific herring live up to fifteen years (Barton and Wespestad, 1980; Stokesbury et al., 1999). Therefore, age-3 and older fish are included in the model and a plus group is used, for herring of ages 9 and older, to minimize the effects of ageing error (Funk and Sandonne, 1990). Estimated parameters include the numbers at age for every age-class in the first year (1980) and recruitment at age 3 in the remaining 32 years (Table indices 1.2.21–23 and Table 1.3). Unlike the current ADF&G model, the Bayesian model did not use the Ricker stock recruit relationship to keep estimated age-3 abundance positive, but instead used broad, positive uniform priors on the log-scale for recruitment, consistent with the assumption that herring recruitment is log-normally distributed. The remaining age-classes in 1980 were also estimated using broad, positive uniform priors in log-space. These parameters, along with fishery removals, natural mortality and gear selectivity parameters, supply information to complete a population matrix of pre-fishery total abundances for ages 3 to 9+ across years 1980–2012 (Eqs 1.4.8–9).

Each model year begins at the start of the spawning season in the spring, and fishing events occur in the spring and fall. For use in the pre-fishery abundance calculations, it was necessary to derive annual purse seine catch in millions (Eq. 1.4.1) using observed purse-seine annual yield (mt), purse-seine proportion of catch at-age, and weight-at-age (mt) data (Table indices 1.1.11, 1.1.8, and 1.1.7, respectively). In common with the spring fisheries, all surveys are conducted during the pre-spawn and spawning period when herring form large aggregations and are easier to sample. The exception is the diver survey, which measures numbers of total herring eggs deposited in the Sound and therefore must be conducted after spawning is complete.

Two non-fishery instantaneous mortality rates that are age-specific and time-independent were used to model biomass losses not attributed to fishing (predation, starvation, disease, etc.). One of these rates applies to herring of ages 3–8 and was assumed to remain constant at the value of 0.25 yr^{-1} prior to 1992 (Table index 1.2.1 and Table 1.3) since 0.25 yr^{-1} is thought to be the lowest realistic rate of average, instantaneous mortality for Pacific herring (Quinn et al., 2001). The other non-fishery mortality rate applies to fish in the plus-group and was estimated using a uniform prior with bounds of (0.30 yr^{-1}, 2.00 yr^{-1}) (Table index 1.2.2), where the lower bound reflects the assumption that instantaneous mortality for the oldest fish in the population is higher than that of younger fish due to senescence, and the upper bound for this mortality was chosen to result in a sufficiently broad prior distribution. These rates are used to discretely model cumulative herring losses between fishing seasons (Eqs 1.4.3–6) since fishing events occur in two periods and last for only a short time (Funk and Sandonne, 1990).

There is evidence of recent high levels of disease significantly affecting Pacific herring abundance (Marty et al., 2003). Thus, additional mortality due to increased disease is added to age 3–8 herring mortality starting in model-year 1992 (Eqs 1.4.4–5). The assumed functional form incorporates the assumption that disease mortality and survival are non-linearly related (Hulson et al., 2008; Marty et al., 2010). VHSV infection is assumed to increase the rate of mortality for fish of ages 3–4, and increased mortality due to *I. hoferi* infection impacts fish of ages 5–8. Estimated parameters scale each disease-infection index to mortality and can take on non-negative values with broad uniform priors (Table indices 1.2.3–5). This study introduced separate *I. hoferi* scalars over 1994–2006 and 2007–2012 (Table indices 1.2.4–5) to capture the potential difference in the relationship between the index and true mortality resulting from a change in detection method from histopathology to tissue explant culture.

One key assumption in the model is that additional mortality from disease began in 1992 even though the disease index begins in 1994. Thus, additional disease mortality in 1993 was estimated using broad, non-negative uniform priors (Table indices 1.2.6–7), and mortality in the latter half of 1992 is set equal to that for 1993. This is important since the estimated additional mortality in 1992 and 1993 allows the model to account for the collapse in biomass that is evident in the milt and hydroacoustic indices of abundance (Marty et al., 2010; Hulson et al., 2008). Therefore, another way of thinking about “disease mortality” in these two years is that it is an estimate of the additional mortality from all sources (disease, predation, competition, oil-spill effects) required to explain the sharp decline in biomass over 1992–1993.

Gear selectivity is assumed to be a logistic function of age *a,* with two parameters, α_v and β_v (Eq. 1.4.7). Modeled in this way, gear selectivity is interpreted as the proportion of age-a fish that will be caught by the purse-seine fishery per unit fishing mortality. Herring recruit into the model population at age-3 and this is the first age that is vulnerable to fishing in the model, therefore parameter α_v, which is the age at which 50% selectivity occurs, was estimated using a uniform prior with bounds (3, 5) where the upper bound was chosen to yield a sufficiently wide interval. Parameter β_v is the slope at 50% selectivity, and a uniform prior with bounds (1, 7) was assigned to this parameter. It was chosen to be reasonably broad.

Herring in Prince William Sound first spawn between the ages of three and five (Barton and Wespestad, 1980). Therefore, maturity proportions for age-3 and age-4 herring were estimated, in lieu of maturity-at-age data, using the assumption that age-5 herring were fully mature (Table indices 1.2.17–20). Previous studies indicate that maturity schedules for age-3 and -4 herring in PWS changed after 1997 (Hulson et al., 2008), thus different sets of maturity parameters were estimated for the two periods. The proportion of age-3 fish that are mature is forced to be less than the proportion of such age-4 fish. Specifically, the maturity of age-3 fish is estimated as a fraction of the age-4 maturity parameter that is bounded between (0.00, 0.75) (Table index 1.2.17). The proportion of age-4 herring that are mature after 1997 was held constant at 0.9 to ensure model convergence; without this assumption, this parameter was stuck at its upper bound, affecting differentiability (Table index 1.2.20). These maturity parameters and empirical weight-at-age relationships were used to transform total abundance (Eqs 1.4.8–9) into spawning biomass (Eqs 1.4.12–13).

1.2.5 *Model formulation: likelihood components and expressions*

The original ASA model minimized sum of squares for the fit of the model to the data. In the updated model, we use likelihoods so that statistical weights can be assigned to each dataset automatically. Six likelihoods relate model estimates to the observed age-compositions from the purse-seine catch and herring-spawn survey data, egg deposition estimates, two hydroacoustic biomass indices, and mile-days of milt index (Table 1.5).

Both purse-seine fishery and ADF&G herring-spawn survey age-compositions were estimated from gear-selectivity and the ratio of numbers-at-age in year y to total numbers in year y (Eqs 1.4.19–20). I assumed a multinomial distribution for the proportions-at-age from the purse-seine fishery and the ADF&G herring-spawn survey age-compositions (Eqs 1.5.2–3). Due to schooling behavior of herring and gear-selectivity, recorded sample sizes need to be decreased to levels that reflect the actual variance contained in the sample. The effective sample size for each series i for year y ($Z_{i,y}$) was estimated using a modified version of the iterative reweighting procedure (McAllister and Ianelli, 1997):

$$Z'_{i,y} = \sum_{a=3}^{9+} \frac{\hat{\theta}_{i,y,a}(1 - \hat{\theta}_{i,y,a})}{\theta_{i,y,a} - \hat{\theta}_{i,y,a}},$$

which takes advantage of a ratio of observed to estimated age-compositions from series i across years and iteratively estimates sample sizes until the process converges to a final set. This set of effective sample sizes was then supplied to the ASA model as input for the MCMC runs (see below). We modified this approach by using the harmonic mean (across years) of the ratio of estimated to recorded sample size in each iteration to reduce the recorded sample size for use in the next iteration of the reweighting algorithm, as recommended by Stewart et al. (2014).

Naturally spawned eggs were estimated from data on the numbers of eggs per female of age a multiplied by the numbers-at-age of female spawners for year y given by the post-fishery spawning abundance estimates and proportion-female data (Eq. 1.4.21). A key assumption was that the egg deposition data were used as an absolute index of abundance. Predictions were tuned to observed egg deposition using the assumption of log-normally distributed errors (Eq. 1.5.4) with annual coefficient of variation (CV) of $\sigma_{E,y}$ (Eq. 1.5.5).

The egg deposition data and the PWSSC hydroacoustic data include sample annual 95% confidence intervals. The methods outlined in Buckland (1992) were used to derive sample CVs for the data (Table indices 1.1.13 and 1.1.16), using the assumption that the CV of a log-normally distributed random variable asymptotically approaches the standard deviation of the logarithm of that variable. The model utilizes these survey-derived CVs, along with estimated additional error (Table indices 1.2.13–14), to characterize total uncertainty (Eqs 1.5.5–8).

Model biomass was multiplied by an estimated scaling factor, q_1, to compare it to the ADF&G hydroacoustic biomass estimates (Eq. 1.4.17) of mature and immature herring of ages 3 and older in the Sound before the spring catches are removed. This scaling factor q_1 was estimated using a broad, uniform prior (Table index 1.2.10), and a lognormal distribution was assumed for the sampling distribution for the hydroacoustic data (Eq. 1.5.6). Since no estimates of sample variance were provided, total uncertainty in the acoustic biomass was estimated as a year-independent CV (Table index 1.2.11) to represent model, process, and observed variance (Francis, 2011).

Model biomass was multiplied by a separate estimated scaling factor, q_2, to compare it to the PWSSC hydroacoustic biomass estimates (Eq. 1.4.18). The scaling factor q_2 was estimated using a log-link and a broad, uniform prior (Table index 1.2.12) under the assumption of log-normally distributed measurement errors (Eq. 1.5.7). As in the egg deposition component (discussed above), survey-derived sample CVs were used, along with estimated additional error (Table index 1.2.13), to characterize total uncertainty (Eq. 1.5.8).

Predicted mile-days of milt were estimated using the ratio of male post-fishery spawning biomass to the parameter q_T, which represents the tonnage of male biomass required to produce a mile-day of herring spawn in log-space (Table index 1.2.15). A lognormal likelihood was assumed for the model fits to the milt index (Eq. 1.5.9), and the bounds for the q_t parameter were chosen to be sufficiently wide as to effectively implement an uninformative prior in log-space. As with the ADF&G acoustic estimates, the CV for milt was estimated (Table index 1.2.16) since no estimates of sample variance were provided.

In order to perform a comparison of variances between the current ADF&G and the Bayesian models, the assumed values used by ADF&G to weight each sum of squares term for the three indices of biomass were used to derive the implied lognormal standard errors that would result in the assumed weights (Francis, 2011). To further facilitate comparison of the assumed variance in the egg deposition data between the two models, the median across survey years of the total egg deposition CV (Eq. 1.5.5) was used in the comparison.

1.2.6 *Projected pre-fishery run biomass in the next year of the model*

For every saved set of parameters from the MCMC chain, pre-fishery run biomass was projected for the next year of the model, 2013. This biomass forecast is the primary management metric used to regulate herring harvest rates in the Sound and refers to the spawning biomass at the start of the spawning season, which is the expected biomass available to the spring harvesters. We extended the population-at-age matrix by one year to obtain a projection of age-4 and older abundance in 2013 and used the mean log-recruitment from the previous 10 years to estimate projected recruitment. The choice of using the mean of recruitment under-estimates uncertainty in the forecast by ignoring variation about the mean, but the running average of recruitment over the latter twenty years of the modeling horizon (1992–2012) has low variability, so projection was robust to longer or shorter intervals for the recruitment mean (results not shown). This abundance-at-age projection was converted to biomass-at-age using the arithmetic mean weight of each cohort from the previous five years and the estimated maturity parameters.

1.2.7 *Implementation and convergence tests*

The ASA model is implemented in an Excel spreadsheet. The Bayesian ASA model used in this study employed the version of the Metropolis-Hastings MCMC sampler included in the ADMB program (Chib and Greenberg, 1995; Fournier et al., 2012) to construct a single chain of eleven-million iterations with a burn-in of 10%. Convergence was reasonably accepted for all parameters using auto-correlation factor <0.10 and Geweke diagnostic statistic z <1.96, where the effective sample size for estimating the mean of each parameter across the MCMC chain, corrected for autocorrelation, was large enough to compare the Geweke statistic to the standard normal distribution.

1.2.8 *Sensitivity tests and retrospective analysis*

A test of the modeling assumption that instantaneous, time-invariant background mortality of herring of ages 3–8 is equal to 0.25 yr^{-1} (Table index 1.2.1) was performed by running the Bayesian ASA Model using values of 0.15 yr^{-1} and 0.35 yr^{-1} for this parameter and comparing results. Evidence for a retrospective pattern was investigated by performing five retrospective runs of the Bayesian ASA Model starting in the same year (1980) and progressing for one fewer year each time (Mohn, 1999). In each retrospective run, the final year's data are "peeled away" and the resulting forecast for year *n* using data ending in year *n-1* is compared to the model estimate of year *n* using data ending in year *n* to reveal systematic over or under-estimation by the assessment model. The degree of retrospective bias for the forecast biomass was quantified using Mohn's ρ, which is the average (across retrospective "peels") relative difference between the most recent estimate from a retrospective run and that from the "reference model" in the same year $B_{2013-y,\text{ref}}$, which was the median pre-fishery run biomass from the current model using the entire data set (Hurtado-Ferro et al., 2014; Mohn, 1999). Therefore, Mohn's ρ for this study is defined as:

$$\rho = \overline{\left(\frac{B_{2013-y,p} - B_{2013-y,\mathrm{ref}}}{B_{2013-y,\mathrm{ref}}}\right)}$$

where the first subscript tracks the number of years into the past of the most recent estimate from the retrospective run of a given "peel" p.

1.3 RESULTS

1.3.1 *Comparison of data weighting*

The data weighting values used by ADF&G, expressed as lognormal standard errors, are included in the 95% credible intervals for the acoustic and milt CVs (Table 1.6). The implied standard errors used in the ADF&G ASA model for the egg deposition and milt data are approximately 10% larger than the median estimated CV for these data types, but the implied standard error used by ADF&G is approximately 30% smaller than the median CV for milt estimated by the Bayesian model (Table 1.6).

1.3.2 *Comparison of Bayesian posteriors to indices of biomass*

This section compares posterior medians and 95% posterior predictive intervals to the observed means and 95% confidence intervals (CI) of each time series used in model fitting. The 95% CI for each time series were derived using the posterior median (in the case of estimated) or assumed coefficients of variation (CV) listed in Table 1.2 (Buckland, 1992).

Milt mile-days were well fitted by the Bayesian model given the 95% posterior predictive intervals (Fig. 1.5, panel A). The total estimated CV for the milt mile-days had a median and 95% credible interval of 0.33 (0.25, 0.44) (Table 1.6) and the estimated tonnage of male biomass required to produce a mile-day of herring spawn was 322.58 (95% interval 248,710: 421,780) (Table index 1.2.15). The model estimate of egg deposition in 1989 was below the 95% posterior predictive interval, but the predictive intervals for the remaining years encompassed the data points and therefore were reasonably good (Fig. 1.5, panel B).

ADF&G acoustic biomass was predicted using total pre-fishery biomass multiplied by the estimated ADF&G acoustic scalar (Table index 1.2.10). The 95% posterior predictive intervals contained all five mean observations from the ADF&G acoustic biomass survey; meaning fits were reasonably good (Fig. 1.5, panel C). The 95% probability interval for the log-link scalar included zero (-0.72, 0.03), which translates to an interval of (0.49, 1.03) and a median of $\exp(-0.34) = 0.71$ on the natural scale. Since the interval contains zero, the posterior includes the possibility that the ADF&G hydroacoustic survey is, on average, an unbiased estimate of the pre-fishery total biomass in absolute terms (Eq. 1.4.11). The PWSSC acoustic biomass was estimated using total pre-fishery biomass multiplied by the estimated PWSSC acoustic scalar (Table index 1.2.12). The 95% posterior predictive intervals contained all twenty mean observations from the PWSSC hydroacoustic biomass survey (Fig. 1.5, panel D). The 95% probability interval for the log-link scalar, translated to the natural scale, is (0.56, 0.96) with a median of $\exp(-0.30) = 0.74$, meaning the PWSSC hydroacoustic survey on average underestimates pre-fishery total biomass (Eq. 1.4.11)

1.3.3 *Comparison of Bayesian posteriors to age-composition data, and recruitment*

Strong cohort signals, which can be clearly identified across consecutive years, are present and consistent in both sets of estimated age-compositions (Fig. 1.6). The largest estimated cohorts were spawned in 1976, 1980, 1984, 1988, and 1999, and clearly match the strongest cohort signals identified in the seine and herring-spawn age-composition data (Fig. 1.6, Fig. 1.7 panel A) for cohorts born in 1980, 1984, 1988, and 1999. Similar signals are seen in the estimates of age-3 fish (Fig. 1.7 panel A), where the largest events occurred in 1983 (median of 446 million fish), 1987 (median of 1,234 million fish), and 1991 (median of 840 million fish) and the smallest recruitment events occurred in 1999 and 2011 with estimated median recruitment of fewer than 10 million fish (Table 1.3).

1.3.4 *Pre-fishery run biomass*

The population collapse of 1992-93 is evident in the estimated biomass trajectory where the highest estimates in the post-collapse period are less than the lowest estimated biomass from the pre-collapse period (Fig. 1.7, panel B). Biomass was highest in 1988–89 (median and 95% intervals of 120,750 (93,920, 161,130) and 124,900 (98,310, 162,760), respectively), and lowest from 1999–2002 (all 95% intervals in this period are contained by a range of 8,740 mt to 20,640 mt) and again from 2005–7 (all 95% intervals contained within 9,570 mt and 20,540 mt). The posterior median of the final year biomass (2013) was estimated to be 19,410 mt with a 95% credible interval of (12,150 mt, 31,740 mt) (Fig. 1.7, panel C and Table 1.3, third column).

1.3.5 *What is the probability that biomass will fall below the lower regulatory threshold?*

There was a 0.54 probability that the biomass forecast in 2013 was below the lower regulatory threshold of 19,555 mt (Table 1.3). The probability was zero that pre-fishery run biomass was below the lower regulatory threshold from 1980–93 (Table 1.3; Fig. 1.7, panel B). For the post-collapse period, there was a probability of 0.00 that biomass in 1997 was below the lower regulatory threshold, which corresponds to a sharp increase in both the PWSSC acoustic biomass (observed value of 37,400 mt) and the milt index (observed value of 64.30 mile-days) in that year. Conversely, the probability that herring biomass was low enough to warrant closure of the commercial fishery was > 0.90 in 1999–2007, except for 2003–4.

1.3.6 *Exploitation rate*

Exploitation rate was defined as total catch (converted to mt using the empirical weight-at-age matrix (2.1.7)) divided by the 95% credible intervals for pre-fishery run biomass. Median exploitation for all years was estimated to be less than or equal to 0.22, and the highest median exploitation rates occurred in 1981 and 1992 (Fig. 1.7, panel D). In 1989, 1996, and 1999, access was limited to either the food/bait or spawn-on-kelp fisheries; therefore the lowest exploitation rate resulting from a full commercial harvest occurred in 1983 (median and 95% interval: 0.05 (0.03, 0.06)) (Table 1.3).

1.3.7 *Selectivity and life history parameters*

The Bayesian model provided estimates of the age and instantaneous rate of change at 50% selectivity using purse-seine gear, and proportions of fish at ages 3 and 4 that are mature in the first and second regime (Table 1.2, Table indices 1.2.8–9 and 1.2.17–20; Fig. 1.9). Multiple parameters used to model total instantaneous mortality for herring of ages 3–9^+ were also estimated (Table 1.2, Table indices 1.2.2–7). Total instantaneous mortality (including additional mortality from VHSV) for herring of ages 3–4 was highest in 1992–1994 and in 1998, and was below 0.40 yr^{-1} in the remaining years (Table 1.3; Fig. 1.10). The rate of total instantaneous mortality (including additional mortality from *I. hoferi)* for herring of ages 5–8 was highest in

1992–1993 and 2003 (Table 1.3; Fig. 1.10), and was below 0.50 yr^{-1} for the remaining years (Table 1.3; Fig. 1.9).

1.3.8 *Sensitivity analysis and retrospective pattern*

Increasing the fixed rate of background mortality (Table index 1.2.1) from the assumed value of 0.25 yr^{-1} to 0.35 yr^{-1} led to an increase in the scale of recruitment for all years, but also resulted in an increase in the scale of the estimated biomass for all years such that the median estimated forecast biomass was approximately 2,000 mt greater than that using 0.25 yr^{-1} (median and 95% interval: 21,210 (13,250, 37,040)) (Fig. 1.11). Decreasing the rate of background mortality to 0.15 yr^{-1} led to a decrease in the scale of recruitment for all years and resulted in a decrease in the scale of the estimated biomass for all years such that the median estimated forecast biomass was approximately 2,000 mt less (median and 95% interval: 17,450 (10,610, 29,580)) (Fig. 1.12). The retrospective analysis revealed systematic underestimation in the posterior medians of pre-fishery run biomass in the most recent five years compared to the ASA Model using the entire duration of available data (Fig. 1.13) and the average Mohn's ρ for these peels was -0.15.

1.4 DISCUSSION

There is a 95% probability that pre-fishery herring biomass in 2013 was within the range 12,150 mt and 31,740 mt, with a median value of 19,410 mt. This study also demonstrated several benefits of adopting the Bayesian version of the PWS herring ASA Model, namely the ability to work with a conceptually simple definition of uncertainty via credible intervals, and the assignment of probabilities associated with alternative states of the population, such as the chance pre-fishery run biomass is below the lower regulatory threshold, which are useful for risk analysis. Furthermore, the Bayesian ASA model lacks evidence of a retrospective pattern in the forecast, but is sensitive to assumed values of baseline non-fishery mortality.

1.4.1 *Comparison of Bayesian estimates of pre-fishery run biomass and the current ADF&G model*

ADF&G's forecast for 2013 was 23,673 mt, which is above the lower regulatory threshold of 19,958 mt. The ADF&G estimate is within the 95% credibility interval from the Bayesian model

(12,150 mt, 31,740 mt) (Fig. 1.5, panel C; Table 1.3, third column), and is approximately 24% higher than the Bayesian posterior median. The principal structural difference between the model developed in this study and ADF&G's current herring model is the way uncertainty is modeled and incorporated into the estimation process, and the statistical basis in the Bayesian model for weighting of different data sources. One of the most important reasons to conduct fisheries stock assessment is to be able to evaluate the consequences of alternative management actions (Punt and Hilborn, 1997). Therefore, one compelling reason to adopt the Bayesian model is to facilitate this type of management decision analysis using objective and intuitive "weights" of population states under different management scenarios in the form of posterior probabilities.

The transformed ADF&G weights, expressed as lognormal standard errors, were similar to the medians from the Bayesian model for ADF&G acoustic biomass and the egg deposition CVs. However, the 95% intervals for the CV of the ADF&G acoustic biomass estimates include values that are half that and close to double the assumed ADF&G error. Furthermore, the median CV on milt estimated by the Bayesian model is approximately 50% higher than the assumed ADF&G value, so the Bayesian model does involve more uncertainty, on average, in the milt data than the weighted least squares model.

1.4.2 *Benefits and implications of using a Bayesian PWS herring assessment model*

Adopting a Bayesian assessment model will require a revision of the regulations set by ADF&G to manage the PWS herring population. As previously mentioned, ADF&G compares the point estimate forecast of pre-fishery run biomass from their weighted least squares assessment model to the lower and upper regulatory thresholds to set the season's harvest rate (Sheridan et al., 2013). However, the Bayesian forecast, in the form of a probability distribution of pre-fishery run biomass, introduces important management questions: Should the median or the mean of the distribution be used in place of the point estimate, and how should uncertainty in the forecast in the form of credible intervals be used to set harvest rates?

Regulations that compare only a measure of central tendency (the mean, median, or mode) of the posterior distribution to a limit do not utilize the complete Bayesian model output. The mean and mode of the posterior distribution of a management metric, such as the forecast of biomass, may fall into any interval of the distribution leading to regulations that have shifting associated probabilities from year to year. The posterior median represents the 50th percentile of the distribution, and is therefore robust to the aforementioned limitation, but as with the previous two measures, the median contains no information on the variance associated with the forecast biomass.

Probabilistic decision rules that take into account the posterior distribution of forecast biomass are possible with a Bayesian assessment model. A menagerie of stochastic assessment regulations exist and there are methods to help managers determine or design the best performing rule given the specific ecological, economic, or conservation objectives tied to the status of their stock. One possible type of decision rule may require a specific probability (or greater) that the stock is above a certain limit for fishing to continue. For example, Kurota et al., (2010) outlines how southern Bluefin tuna is managed by requiring a 70% probability that the biomass is above a threshold value. These managers considered the tradeoffs and decided that a 70% probability of the stock being above the fishable biomass limit was appropriate. Hypothetically, a more conservative approach would require a 90% probability, while a riskier regulation would use a 30% probability. The modifiers “conservative” and “risky” in this example only refer to the increase or decrease in associated probability of the true biomass being above the regulatory limit, respectively.

Determining which credible intervals or cumulative probabilities are best for managing PWS herring (60% or 75% or 90%, etc.) involves trade-offs with respect to conservation goals and ecosystem balance, as well as economic concerns, among others, and any recommendations should be made after conducting a risk analysis and/or management strategy evaluation (MSE) using the goals of the agency and the community. In the interim between adopting the Bayesian ASA Model and completing an MSE, a decision rule may need to be adopted and, given the particular importance of this population, my recommendation would be to adopt a conservative rule that requires an estimated 100% probability that the stock is above the lower threshold of 19,958 mt for two consecutive years for a limited harvest rate to be set. This recommendation first assumes that the lower limit adopted by ADF&G in 1995 (Botz et al., 2010) continues to represent the true amount of biomass required to maintain the herring stock and the Prince William Sound ecosystem, for which it is a vital resource. Furthermore, it considers the variability of herring recruitment from year to year and assumes that if herring biomass was sufficiently high last year to satisfy the opportunistic needs of the ecosystem and remain above the limit this year, then there is enough to sustain a limited catch in the current year. Managing this population using a fully risk-averse rule (0% probability biomass is below the limit) relies on the ability of the Bayesian model to integrate over all known sources of uncertainty, but admits the possibility that other driving factors, not yet integrated into the model, may impact the variability of herring biomass and that managers should be as cautious as possible.

Other management agencies also use regulatory rules that extend beyond a single year. For example, the maximum constant yield (MCY) used in New Zealand states that no management regulation should allow the stock to drop below 20% of virgin biomass over a specific time horizon with greater than 10% probability (Francis, 1992). Therefore, in addition to the recommendation to test for appropriate widths of credible intervals, a management strategy evaluation comparing regulation rules involving a combination of risk over a range of horizons would also provide vital information for PWS herring management going forward.

1.4.3 *Model sensitivities*

The conflict between the milt and egg deposition data is in the years 1988–89 and 1991–92 (Fig. 1.3) and it is in these years that the fit to the milt and egg deposition data are degraded – this is because the model splits the differences between the conflicting trends in the egg deposition and milt data by minimizing errors to both fits. The result is that milt is underestimated when egg deposition is overestimated in 1988 and vice-versa in 1991.

Previous research on the PWS herring assessment model involved a sensitivity analysis using time-invariant instantaneous mortality rates between (0.35 yr^{-1}, 0.55 yr^{-1}) and concluded the main effect of higher mortality is larger recruitment and vice-versa (Funk and Sandonne, 1990). The sensitivity analysis conducted here on time-invariant background mortality (Table index 1.2.1) reveals a similar correlation between mortality and the scale of the recruitment estimates; higher baseline mortality results in larger recruitment in all years and lower mortality results in lower recruits, on average. However, varying the assumed value for background mortality also had an impact on the biomass forecast; a background mortality rate of 0.35 yr^{-1} would result in four median biomass estimates out of the last six years of the modeling horizon being above 20,000 mt (Fig. 1.10). An important next step would be to explore the sensitivity of the agency's regulatory threshold to the assumed rate of non-fishery mortality.

I used the guidelines suggested by Hurtado-Ferro et al. (2014) to contextualize the results from the retrospective analysis, which state for a short-lived, sardine-like life history, a Mohn's ρ

outside of the interval (-0.22, 0.30) would be of concern, but a Mohn's ρ within the interval may or may not be of concern and may need further investigation. The magnitude and direction of Mohn's ρ is not related to that of the bias in final-year biomass compared to the original assessment, but it may be related to the direction of some recent, true change in the population with respect to some time-invariant parameter of the model (Hurtado-Ferro et al., 2014). Recent changes in mortality, growth, or selectivity have been identified as potential drivers behind a retrospective pattern (Parma, 1993) and a positive retrospective pattern is of more conservation concern than a negative pattern. Therefore, the results for Prince William Sound herring (Mohn's ρ = -0.15) may give evidence of a recent decrease in true background mortality rate or true selectivity, etc. However, the retrospective pattern of the first five peels are less than the model-estimated uncertainty in biomass (Fig. 1.13), and so this study reveals a slight, negative retrospective pattern that is of little concern.

1.4.4 *Next steps to improve the Bayesian PWS herring ASA model*

Now that the Bayesian structure for the PWS herring ASA model has been implemented, conclusions from future research on this stock as well as information from other herring stocks can be easily integrated into the model in the form of informative, or more informative, priors. Candidate parameters that would benefit from informative priors are recruitment, the maturity at age proportions, and background mortality. Other functional forms of the prior distributions can also be investigated, including using an inverse-gamma distribution for the coefficients of variation. The sensitivity tests performed here should be extended to include testing the impact of the assumed values for the proportion of impounded fish that die and the implied scalar for egg deposition data, which currently has a value of 1.00.

Table 1.1 Time series used in the Bayesian ASA model. The first column lists the index number used in the text, the next columns list the data type and units, n_y refers to number of years that data type was collected, and the final column reports the first and last year of collection. Note some series are discontinuous.

Index	**Data type**	**Units**	**Symbol**	**n_y**	**Years**
1.1.1	Gillnet catch-at-age	millions	$C_{2,y,a}$	15	(1980, 1998)
1.1.2	Pound utilization catch-at-age	millions	$C_{3,y,a}$	16	(1980, 1999)
1.1.3	Food/bait catch-at-age	millions	$C_{4,y,a}$	17	(1980, 1998)
1.1.4	Disease index of VHSV prevalence, ages 3–4	percentage	$d_{1,y}$	19	(1994, 2012)
1.1.5	Disease index of *I. hoferi* prevalence, ages 5–8	percentage	$d_{2,y}$	19	(1994, 2012)
1.1.6	Fecundity-at-age	no. of eggs per female	$f_{y,a}$	7	(1984, 1993)
1.1.7	Weight-at-age of spawning herring	mt/million fish	$w_{y,a}$	33	(1980, 2012)
1.1.8	Purse-seine age-composition	proportion	$\Theta_{1,y,a}$	13	(1980, 1998)
1.1.9	Spawner survey age-composition	proportion	$\Theta_{Sp,y,a}$	31	(1982, 2012)
1.1.10	Female spawners	proportion	$\rho_{f,y}$	33	(1980, 2012)
1.1.11	Total annual purse-seine yield	mt	$\Omega_{1,y}$	13	(1980, 1998)
1.1.12	Eggs deposited	trillions	E_y	10	(1984, 1997)
1.1.13	C.V. for eggs deposited		$\sigma_{E,A,y}$	10	(1984, 1997)
1.1.14	ADF&G hydroacoustic survey biomass	mt	$H_{1,y}$	5	(2005, 2009)
1.1.15	PWSSC hydroacoustic survey biomass	mt	$H_{2,y}$	20	(1993, 2012)
1.1.16	C.V. for PWSSC hydroacoustic biomass		$\sigma_{H_{2,A,y}}$	20	(1993, 2012)
1.1.17	Milt	mile/day	T_y	33	(1980, 2012)

Table 1.2 Key model parameter estimates (medians and 95% credible intervals (CI)). All mortality is modeled as instantaneous mortality rates.

Index	Parameters	Symbols	Parameter values Median and 95% interval	Prior
1.2.1	Background mortality, 1980–91, ages 3-	$m_{a \in (3,\ldots,8)}$	0.25	Not estimated
1.2.2	Total mortality, 1980–91, age 9^+	$m_{9^+} = \mu_{9^+}$	0.93 (0.60, 1.30)	$U \sim (0.30, 2.00)$
1.2.3	VHSV disease scalar, ages 3-4	β_1	83.56 (18.18, 159.5)	$U \sim (0.00, 1000)$
1.2.4	*I. hoferi* scalar, ages 5-8, 1994–2006	$\beta_{2,1}$	0.90 (0.33, 1.55)	$U \sim (0.00, 25.00)$
1.2.5	*I. hoferi* scalar, ages 5-8, 2007–12	$\beta_{2,2}$	0.45 (0.03, 1.19)	$U \sim (0.00, 25.00)$
1.2.6	Disease mortality in 1993, VHSV	$m_{1,1993,a} = \mu_{1,1993}$	0.67 (0.16, 1.18)	$U \sim (0.00, 5.00)$
1.2.7	Disease mortality in 1993, *I. hoferi*	$m_{2,1993,a} = \mu_{2,1993}$	0.68 (0.23, 1.09)	$U \sim (0.00, 5.00)$
1.2.8	Purse-seine gear selectivity	α_V	3.78 (3.49, 4.08)	$U \sim (3.00, 5.00)$
1.2.9	Purse-seine gear selectivity	β_v	2.27 (1.60, 3.20)	$U \sim (1.00, 7.00)$
1.2.10	ADF&G acoustic scalar, log-link	q_1	-0.36 (-0.79, 0.07)	$U \sim (-5.00, 5.00)$
1.2.11	ADF&G acoustic biomass CV	σ_{H_1}	0.29 (0.15, 0.55)	$U \sim (0.00, 0.60)$
1.2.12	PWSSC acoustic scalar, log-link	q_2	-0.31 (-0.63, 0.01)	$U \sim (-5.00, 5.00)$
1.2.13	PWSSC acoustic biomass add'l error	$\sigma_{H_2,B}$	0.34 (0.24, 0.52)	$U \sim (0.00, 0.60)$
1.2.14	Egg deposition additional error	$\sigma_{E,B}$	0.40	Not estimated
1.2.15	Milt scalar, log-link	q_T	322.58 (248.71, 421.78)	$U \sim (2.30, 7.00)$
1.2.16	Milt CV	σ_T	0.33 (0.25, 0.44)	$U \sim (0.00, 0.60)$
1.2.17	Proportion mature at age 3, 1980–96	$\nu_3;\ \rho_{M,1,3} = \nu_3 \rho_{M,1,4}$	0.39 (0.28, 0.56)	$U \sim (0.00, 0.75)$
1.2.18	Proportion mature at age 4, 1980–96	$\rho_{M,1,4}$	0.80 (0.62, 0.97)	$U \sim (0.00, 1.00)$
1.2.19	Proportion mature at age 3, 1997–2012	$\rho_{M,2,3}$	0.49 (0.37, 0.66)	$U \sim (0.00, 1.00)$
1.2.20	Proportion mature at age 4, 1997–2012	$\rho_{M,2,4}$	0.90	Not estimated
1.2.21	Recruitment by year (millions), log-link	$\eta_{y,3} = \ln(N_{y,a})$	Table 1.3	$U \sim (0.00, 8.01)$
1.2.22	Age-4 abundance in 1980, log-link	$\eta_{1980,4} = \ln(N_{1980,4})$	6.33 (6.10, 6.57)	$U \sim (0.00, 8.01)$
1.2.23	Age-5 abundance in 1980, log-link	$\eta_{1980,5} = \ln(N_{1980,5})$	4.28 (3.88, 4.68)	$U \sim (0.00, 8.01)$

Table 1.3 Recruitment (median and 95% credible intervals (95% int.)) in millions of age-3 fish, pre-fishery run biomass (median and 95% int.), the probability that pre-fishery run biomass (B) biomass has fallen below the lower regulatory threshold (LRT), exploitation rate (median and 95% int.) defined as total catch in each year (converted to metric tons using the empirical weight-at-age matrix) divided by the 95% biomass trajectories (3rd column), total instantaneous mortality for age 3-4 fish, and total instantaneous mortality for age 5-8 fish.

Year	**Recruitment** median and **95% int.**	**Pre-fishery run biomass (B)** median and **95% int.**	**Prob. B<LRT**	**Exploitation rate** median and **95% int.**	**Age 3-4 total mortality**	**Age 5-8 total mortality**
1980	225.21 (162.41, 308.76)	57.23 (41.86, 76.58)	0.00	0.12 (0.09, 0.17)	0.25	0.25
1981	118.78 (78.67, 173.39)	66.26 (52.80, 85.69)	0.00	0.21 (0.17, 0.27)	0.25	0.25
1982	161.48 (112.27, 230.74)	57.70 (43.89, 77.93)	0.00	0.14 (0.10, 0.18)	0.25	0.25
1983	447.91 (340.35, 594.45)	67.50 (50.36, 91.18)	0.00	0.05 (0.03, 0.06)	0.25	0.25
1984	373.56 (279.42, 497.78)	79.52 (60.26, 106.03)	0.00	0.09 (0.07, 0.12)	0.25	0.25
1985	119.70 (79.73, 179.77)	96.28 (73.94, 126.78)	0.00	0.08 (0.06, 0.11)	0.25	0.25
1986	142.15 (96.18, 207.55)	84.73 (65.96, 110.90)	0.00	0.13 (0.10, 0.17)	0.25	0.25
1987	1237.93 (988.44, 1580.59)	95.60 (73.80, 126.20)	0.00	0.07 (0.05, 0.09)	0.25	0.25
1988	136.76 (94.92, 207.47)	122.03 (94.20, 158.26)	0.00	0.08 (0.06, 0.11)	0.25	0.25
1989	28.43 (18.86, 52.53)	124.99 (98.52, 161.09)	0.00	0.00 (0.00, 0.01)	0.25	0.25
1990	29.67 (11.85, 67.94)	107.57 (85.74, 138.74)	0.00	0.10 (0.08, 0.12)	0.25	0.25
1991	844.99 (566.63, 1287.31)	94.97 (74.38, 123.53)	0.00	0.17 (0.13, 0.22)	0.25	0.25
1992	62.41 (23.70 149.93)	93.74 (69.01, 129.80)	0.00	0.22 (0.16, 0.30)	0.93	0.94
1993	135.11 (65.92, 292.40)	38.83 (29.38, 51.21)	0.00	0.07 (0.05, 0.09)	0.93	0.94
1994	18.25 (6.92, 41.75)	20.49 (14.88, 27.92)	0.44	–	1.09	0.32
1995	94.46 (64.43, 136.20)	18.71 (14.10, 24.71)	0.67	–	0.37	0.36
1996	76.91 (47.93, 121.32)	20.52 (15.52, 27.02)	0.43	0.02 (0.02, 0.03)	0.25	0.36
1997	142.02 (81.34, 246.50)	27.92 (21.65, 36.78)	0.00	0.18 (0.13, 0.23)	0.33	0.35
1998	69.83 (39.03, 124.43)	21.89 (16.53, 29.38)	0.27	0.19 (0.14, 0.25)	0.73	0.39
1999	6.08 (1.4, 18.28)	14.93 (10.48, 21.09)	0.95	0.00 (0.00, 0.00)	0.26	0.36
2000	22.41 (10.23, 42.72)	13.43 (9.46, 18.70)	0.99	–	0.25	0.36
2001	10.27 (3.67, 22.91)	11.91 (8.42, 16.58)	1.00	–	0.26	0.44
2002	210.28 (145.21, 301.15)	14.66 (10.53, 20.46)	0.97	–	0.37	0.40
2003	39.97 (24.69, 63.64)	19.90 (14.48, 27.78)	0.51	–	0.26	0.54
2004	19.03 (10.25, 32.64)	20.33 (14.57, 28.42)	0.46	–	0.26	0.40
2005	26.15 (14.92, 44.22)	15.89 (11.26, 22.08)	0.92	–	0.26	0.40
2006	16.74 (8.77, 29.50)	13.70 (9.50, 19.34)	0.98	–	0.26	0.40
2007	102.28 (68.24, 150.81)	15.36 (10.83, 21.69)	0.94	–	0.25	0.38
2008	94.69 (62.78, 143.46)	21.14 (15.08, 29.76)	0.37	–	0.25	0.33
2009	28.11 (13.26, 55.63)	20.12 (14.36, 28.09)	0.48	–	0.25	0.34
2010	53.48 (22.39, 107.25)	20.58 (14.62, 28.91)	0.43	–	0.27	0.30
2011	9.23 (1.34, 48.71)	18.06 (12.53, 25.58)	0.72	–	0.25	0.32
2012	77.84 (18.94, 226.89)	18.14 (12.13, 26.66)	0.69	–	0.25	0.36
2013	35.29 (23.87, 52.75)	19.41 (12.15, 31.74)	0.54	–	–	–

Table 1.4 Model formulation, first column gives the equation number, the second column gives a description, and the final column gives the mathematical form of the dynamics.

Eq. No	Description	Equation
	Catch, millions of fish	
1.4.1	Estimated total purse-seine catch	$\hat{C}_{1,y} = \frac{\Omega_{1,y}}{\sum_{a \in A}\left(\hat{\Theta}_{1,y,a} w_{y,a}\right)}$
1.4.2	Spring removals, $\rho_k = 0.75$	$\hat{C}_{S,y,a} = \hat{\Theta}_{1,y,a}\hat{C}_{1,y} + C_{2,y,a} + \rho_k C_{3,y,a}$
	Survival, rate	
1.4.3	Half-year survival, 1980–1991, all ages	$S_{y,a}^{.5} = e^{-.5 m_a}$
1.4.4	Half-year survival, 1993–2012, ages 3–4, with VHSV mortality	$S_{y,a}^{.5} = e^{-.5(m_a + \beta_1 d_{1,y})}$
1.4.5	Half-year survival, 1993–2012, ages 5–8, with *I. hoferi* mortality, collection era *x=1,2*	$S_{y,a}^{.5} = e^{-.5(m_a + \beta_{2,x} d_{2,y})}$
1.4.6	Half-year survival, 1993–2012, plus group	$S_{y,9^+}^{.5} = S_{y-1,9^+}^{.5}\left(\frac{S_{y,8}^{.5}}{S_{y-1,8}^{.5}}\right)$
	Selectivity, logistic form	
1.4.7	Purse-seine gear selectivity by age	$V_a = \frac{1}{1 + e^{-\beta_v(a - \alpha_v)}}$
	Abundance, millions of fish	
1.4.8	Pre-fishery total abundance, ages 4–8	$N_{y+1,a+1} = \left[\left(\left(N_{y,a} - \hat{C}_{S,y,a}\right)S_{y,a}^{.5}\right) - C_{4,y,a}\right] S_{y,a}^{.5}$
1.4.9	Pre-fishery total abundance, ages 9$^+$	$N_{y+1,9^+} = \left[\left(\left(N_{y,8} - \hat{C}_{S,y,8}\right)S_{y,8}^{.5}\right) - C_{4,y,8}\right] S_{y,8}^{.5} +$ $\left[\left(N_{y,9^+} - \hat{C}_{S,y,9^+}\right)S_{y,9^+}^{.5} - C_{4,y,9^+}\right]S_{y,9^+}^{.5}$
1.4.10	Post-fishery spawning abundance	$\tilde{N}_{y,a} = \rho_{M,a}\left[N_{y,a} - \left(\hat{\Theta}_{1,y,a}\hat{C}_{1,y} + C_{2,y,a} + C_{3,y,a}\right)\right]$
	Biomass, mt	
1.4.11	Pre-fishery total biomass	$B_y = \sum_{a \in A}\left(N_{y,a} w_{y,a}\right)$
1.4.12	Pre-fishery spawning biomass	$\tilde{B}_y = \sum_{a \in A} \rho_{M,a} N_{y,a} w_{y,a}$
1.4.13	Post-fishery spawning biomass	$\tilde{B}_{post,y} = \sum_{a \in A} \tilde{N}_{y,a} w_{y,a}$
1.4.14	Estimated 2013 pre-fishery run biomass	$B_{2013} = B_{2013,3} + \sum_{a \in A_{-3}} \rho_{M,a} N_{2013,a} \bar{w}_a$
1.4.15	Average weight-at-age over the last 5 years	$\bar{w}_a = \frac{1}{5}\sum_{i=2008}^{2012} w_{i,a}$
1.4.16	Estimated 2013 age-3 biomass	$B_{2013,3} = \rho_{M,2,3}\bar{w}_3 \exp\left(\frac{1}{10}\sum_{i=2003}^{2012}\ln(\eta_i)\right)$
	Estimates used in the likelihood expressions	
1.4.17	Estimated ADF&G hydro-acoustic biomass, mt	$\hat{H}_{1,y} = B_y e^{q_1}$
1.4.18	Estimated PWSSC hydro-acoustic biomass, mt	$\hat{H}_{2,y} = B_y e^{q_2}$
1.4.19	Estimated purse-seine age composition	$\hat{\Theta}_{1,y,a} = \frac{V_a N_{y,a}}{\sum_{a \in A}\left(V_a N_{y,a}\right)}$
1.4.20	Estimated spawning age composition	$\hat{\Theta}_{Sp,y,a} = \frac{\rho_{M,a} N_{y,a}}{\sum_{a \in A}\left(\rho_{M,a} N_{y,a}\right)}$
1.4.21	Estimated naturally spawned eggs, trillions	$\hat{E}_y = 10^{-6}\rho_{f,y}\sum_{a \in A}\left(\tilde{N}_{y,a} f_{y,a}\right) \quad \forall\, y \in Y$
1.4.22	Estimated milt, mile-days	$\hat{T}_y = \frac{\left(1 - \rho_{f,y}\right)\tilde{B}_{post,y}}{e^{q_T}}$

Table 1.5 Components contributing to the negative of the logarithm of the likelihood expression for the Bayesian ASA model.

Eq. No	**Likelihood component**	**Form**
1.5.1	Complete expression	$L = \sum_{i=1}^{6} L_i$
1.5.2	Purse-seine age-composition	$L_1 = -\sum_{y=1980,\dots}^{\sim 1998} \left[Z_{1,y} \sum_{a \in A} \Theta_{1,y,a} \ln\left(\frac{\hat{\Theta}_{1,y,a}}{\Theta_{1,y,a}}\right)\right]$
1.5.3	Spawner survey age-composition	$L_2 = -\sum_{y=1982}^{2012} \left[Z_{Sp,y} \sum_{a \in A} \Theta_{Sp,y,a} \ln\left(\frac{\hat{\Theta}_{Sp,y,a}}{\Theta_{Sp,y,a}}\right)\right]$
1.5.4	Number of eggs deposited	$L_3 = 10 \sum_{y \in Y_E} \left[\ln(\sigma_{E,y}) + \frac{\left(\ln(\hat{E}_y) - \ln(E_y)\right)^2}{2\sigma_{E,y}^2} \right]$
1.5.5	Total variance for L_3	$\sigma_{E,y}^2 = \sigma_{E,A,y}^2 + \sigma_{E,B}^2$
1.5.6	ADF&G hydroacoustic biomass	$L_4 = 5\ln(\sigma_{H_1}) + \frac{1}{2\sigma_{H_1}^2} \sum_{y \in Y_H} \left[\ln(\hat{H}_{1,y}) - \ln(H_{1,y})\right]^2$
1.5.7	PWSSC hydroacoustic biomass	$L_5 = 20 \sum_{y \in Y_H} \left[\ln(\sigma_{H_2,y}) + \frac{\left(\ln(\hat{H}_{2,y}) - \ln(H_{2,y})\right)^2}{2\sigma_{H_2,y}^2} \right]$
1.5.8	Total variance for L_5	$\sigma_{H_2,y}^2 = \sigma_{H_2,A,y}^2 + \sigma_{H_2,B}^2$
1.5.9	Milt mile-days	$L_6 = 33\ln(\sigma_T) + \frac{1}{2\sigma_T^2} \sum_{y \in Y} \left[\ln(\hat{T}_y) - \ln(T_y)\right]^2$

Table 1.6 Comparison of weights between the ADF&G model and the Bayesian model. The first column lists the weights used in the ADF&G weighted least squares model (λ), the second column shows those weights converted to standard errors (σ) (Francis, 2011), and the third column shows the coefficients of variation (CV) used in the Bayesian model for the same data. Median and 95% posterior intervals are shown for the ADF&G hydroacoustic biomass and milt CVs. Sample errors were provided for the years that the egg-deposition survey ran (Table 1.1, row 1.1.13), which were combined with the assumed value of additional error (0.40; Table 1.2, row 1.2.14) to derive total egg deposition CV. Therefore, the median total CV across the survey years is presented for comparison.

	ADF&G model		Bayesian model
Data type	λ	σ	**CV** median and **95% int.**
Eggs deposited	0.25	0.45	0.43
ADF&G hydroacoustic survey biomass	0.50	0.32	0.29 (0.15, 0.55)
Milt	1.00	0.22	0.33 (0.25, 0.44)
PWSSC hydroacoustic survey biomass	-	-	0.35 (0.25, 0.53)

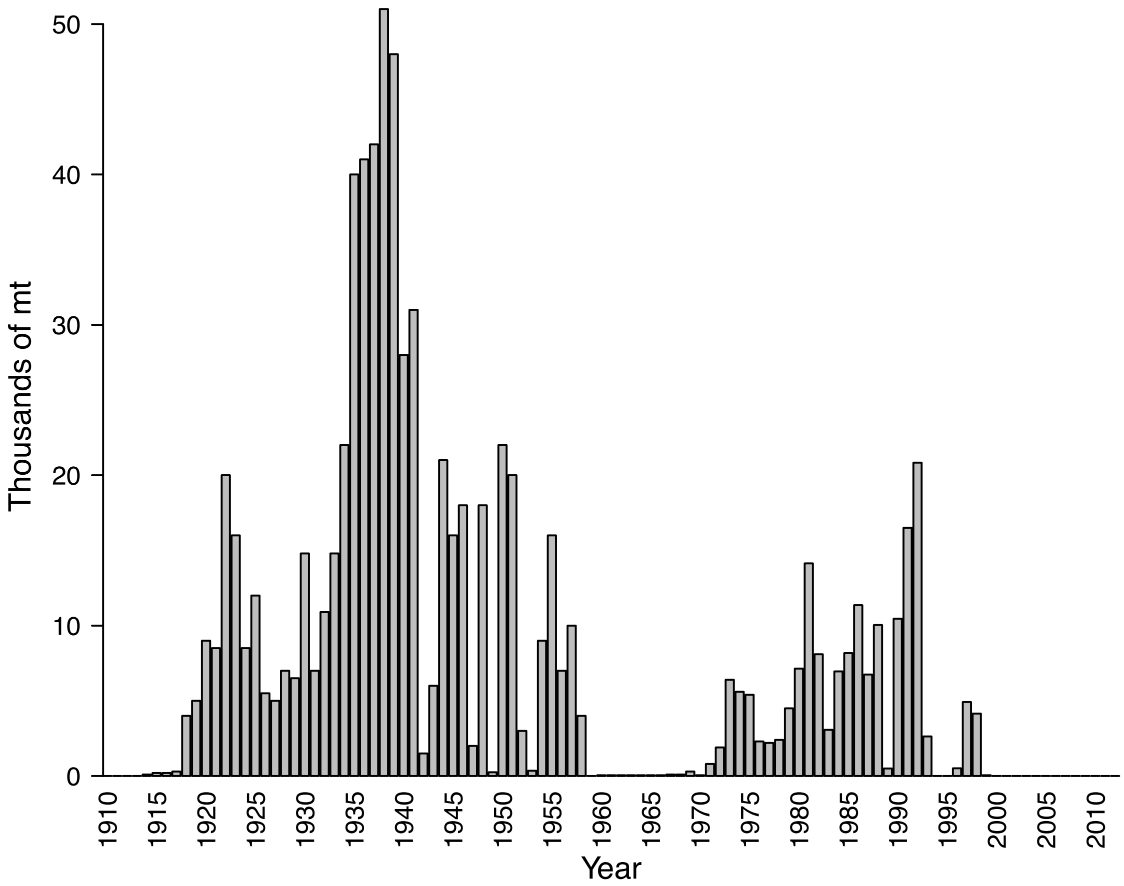

Figure 1.1 Commercial harvests of Pacific herring reported for Prince William Sound, 1914 through 2012 (Funk and Sandonne, 1990).

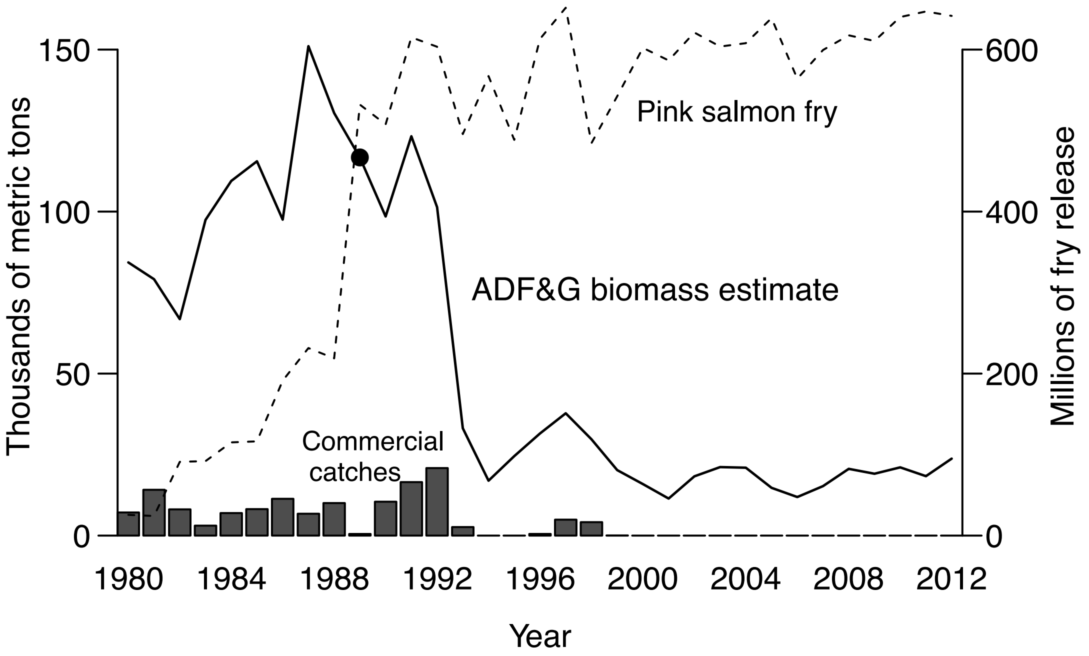

Figure 1.2 ADF&G estimates of pre-fishery run biomass (management metric used to set catch allowance) with bar plots of aggregated commercial catch data from Prince William Sound over the modeling horizon (1980–2012). Both quantities are on the same scale of thousands of metric tons (left axis). The dot denotes 1989, the year of the Exxon-Valdez Oil Spill. The thin black line shows numbers in millions (right axis) of hatchery salmon fry released into rivers that feed into the Sound.

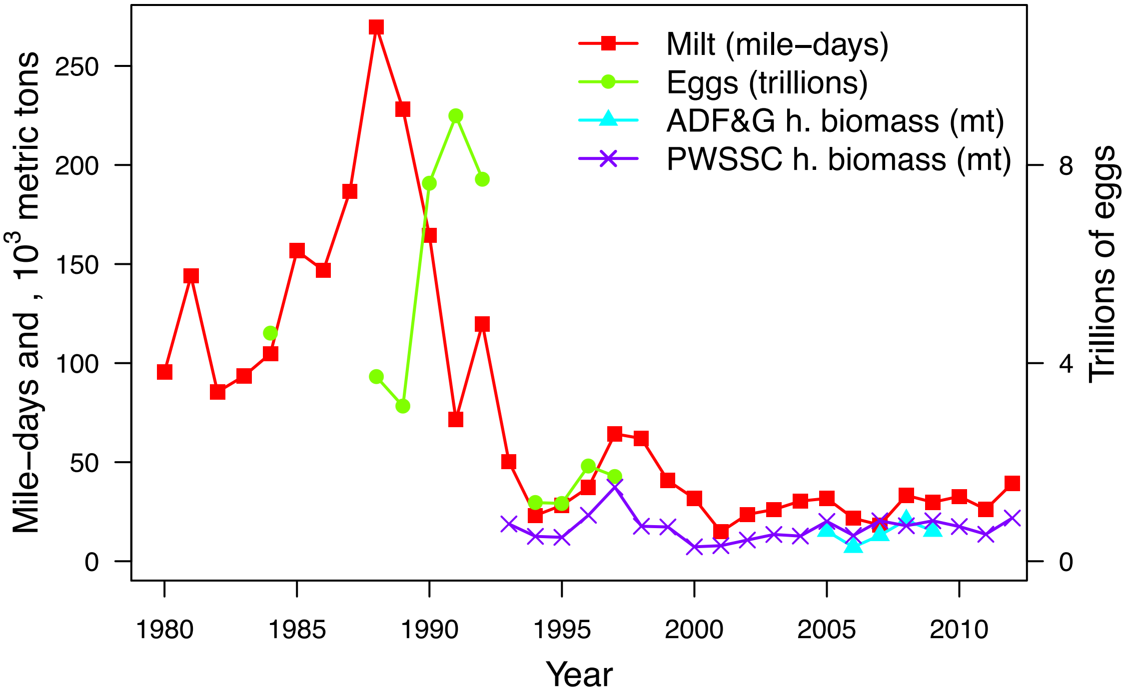

Figure 1.3 The four types of Prince William Sound herring index data (1980–2012) used in the Bayesian age-structured assessment model. Even though mile-days of milt and the two hydroacoustic biomass indices (h. biomass in thousands of metric tons) use different units, they are on the same scale (left axis). Number of eggs deposited is on the right axis.

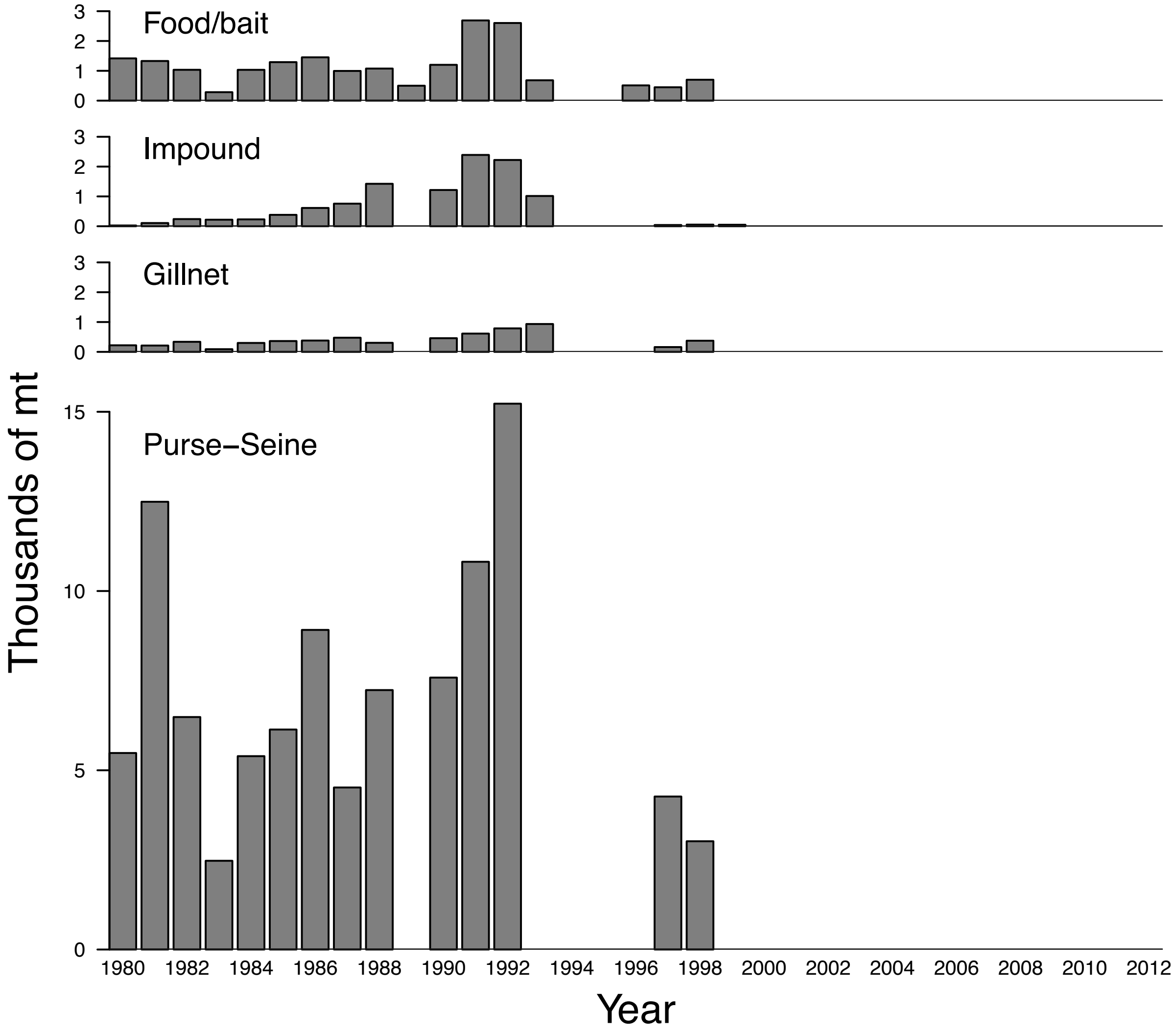

Figure 1.4 The four types of annual catch data, in thousands of metric tons, for Prince William Sound herring used in the Bayesian age-structured assessment model. Data for the three fisheries in the top panels are in the form of numbers of catch-at-age, so these were converted to annual total yield in mt using the weight-at-age (mt) data (1.1.7) for ease of display. Absent bars denote years that fishery did not run; all herring fisheries have been closed since 1999.

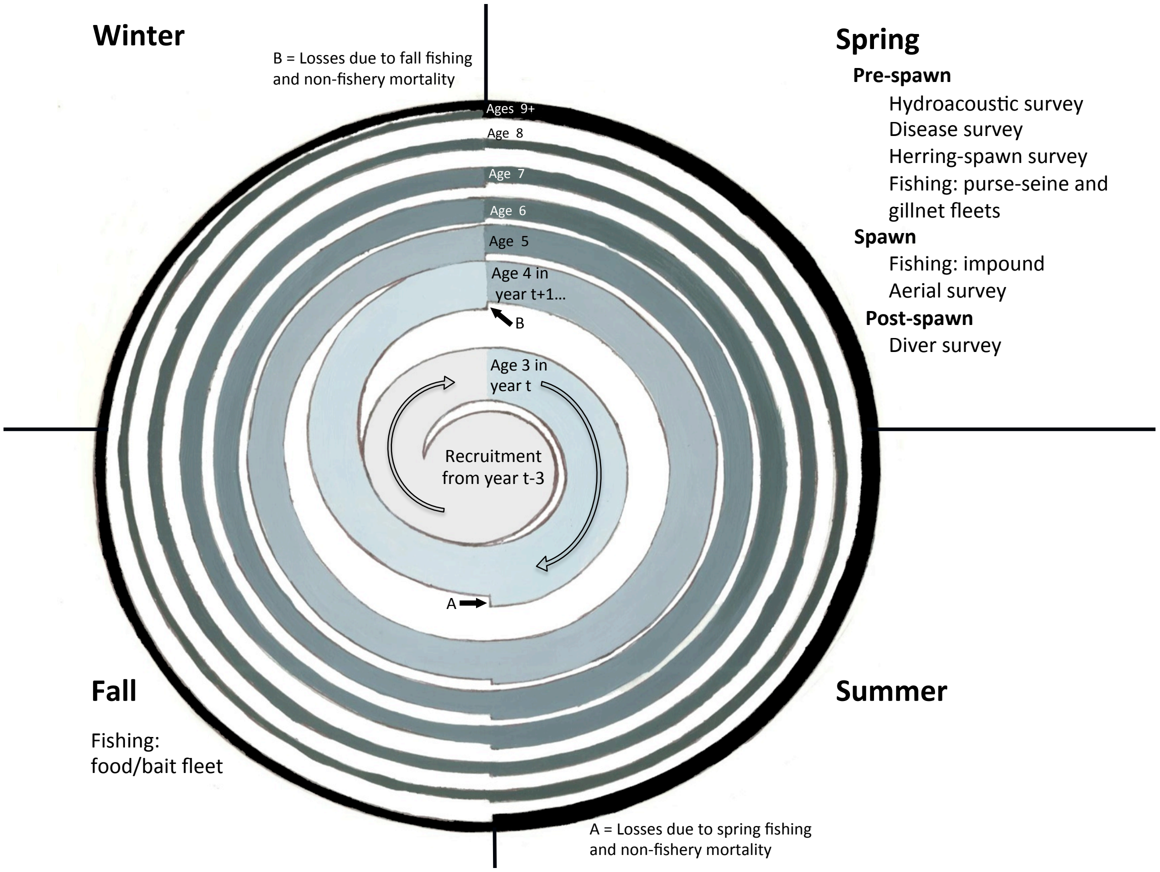

Figure 1.5 A schematic of the seasonal timing of fishing and sampling events included in the assessment model along with a schematic of a single cohort over seven years. Starting in the center of the spiral, the width of each separately colored curl represents the relative size of the cohort at a certain age and lighter colors denote younger ages of the cohort in earlier years. The cohort is reduced by fishery and non-fishery mortality effects (in that order) after the first 6 months (event A) and the last 6 months (event B) of every year before becoming a year older. The plus group is represented as a complete circle with two inputs: herring of age 8 and herring already in the plus group.

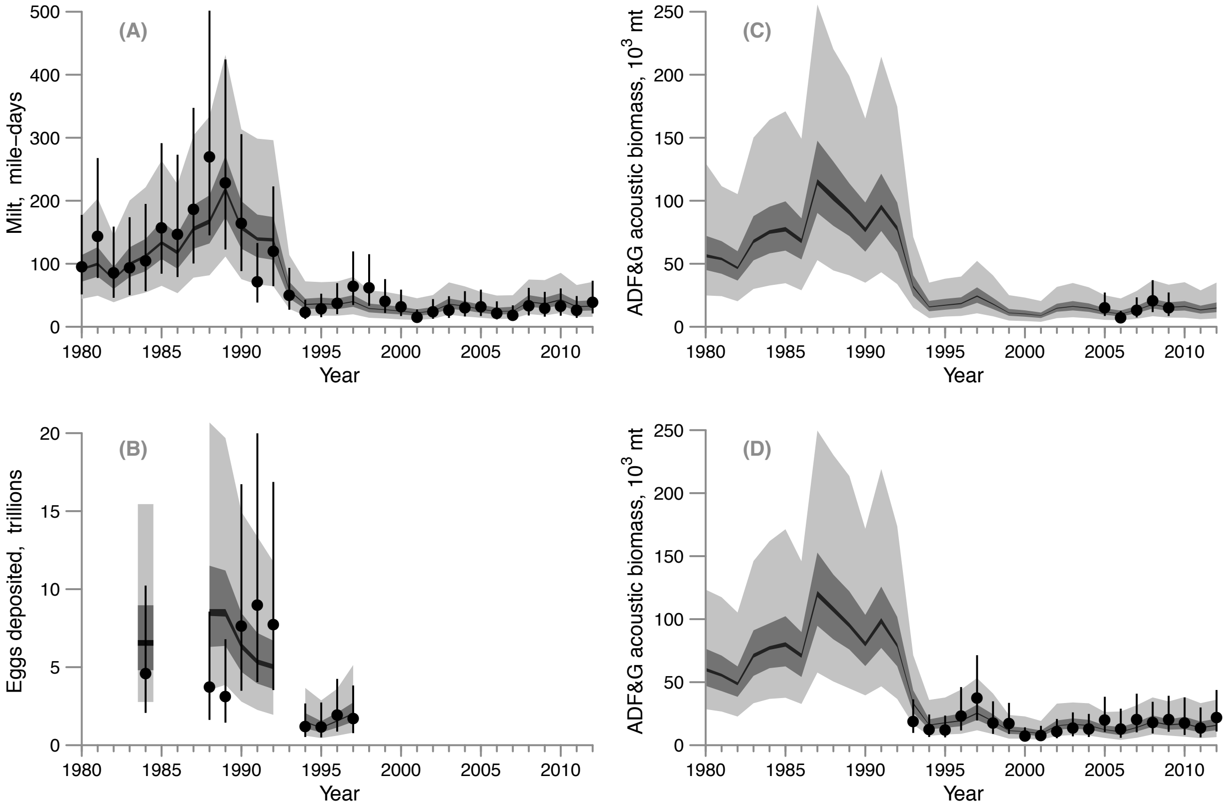

Figure 1.6 Model estimates and the four time series of abundance estimates (1980–2012): (A) mile-days of milt, (B) egg deposition surveys, (C) ADF&G hydroacoustic estimates, and (D) PWSSC hydroacoustic estimates. The solid circles and lines represent the mean and 95% confidence intervals of the data (plus additional variance estimated by the model); the shaded polygons represent the respective posterior predictive intervals (light gray = 95% interval, darker gray = 50% interval, black = 5% interval).

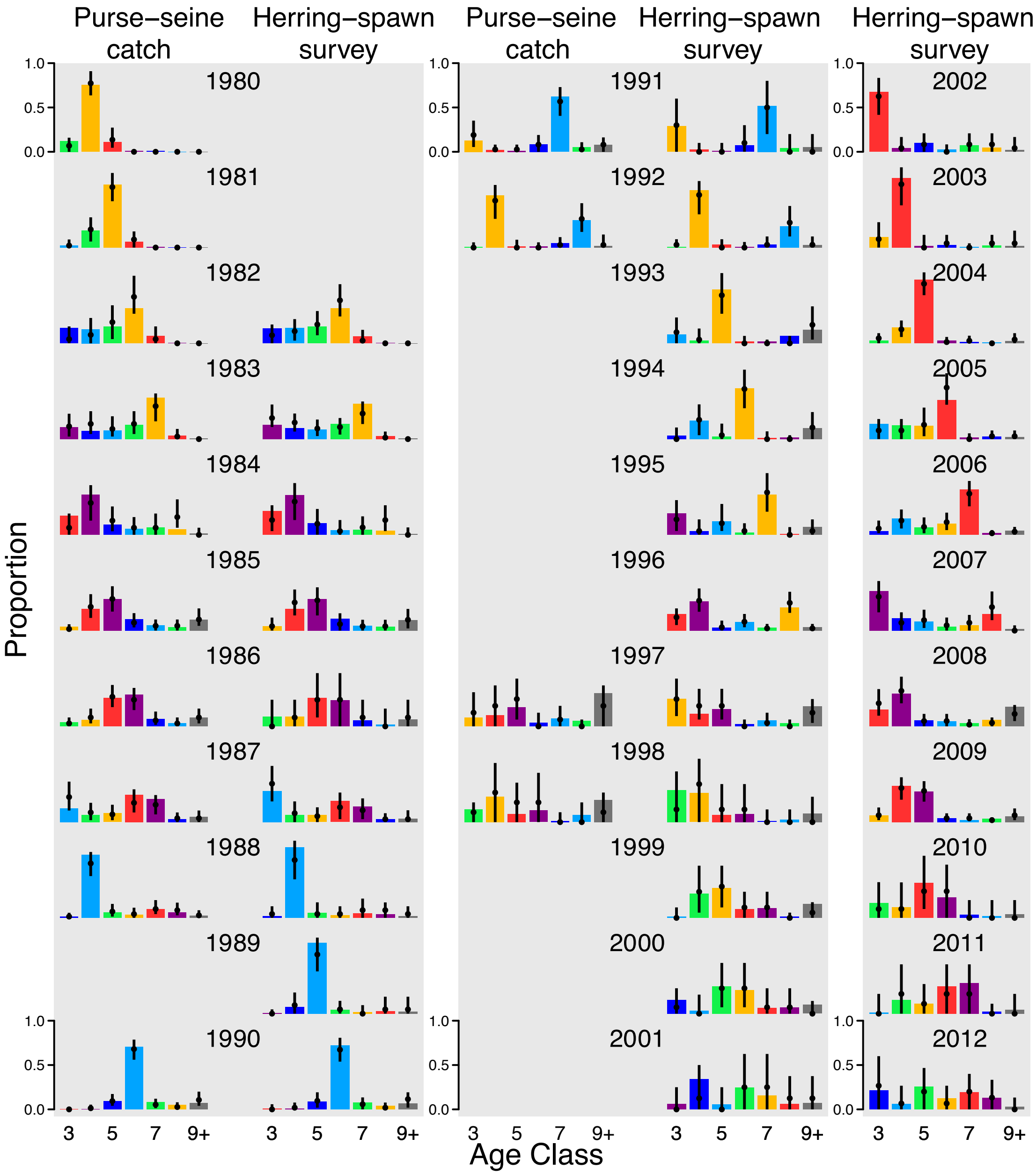

Figure 1.7 Model fits to the two sets of age-composition data used: proportion of catch-at-age from the purse-seine fishery and age-composition proportions from the ADF&G herring-spawn survey. Colored bars denote data, colors track individual cohorts through time, and points show posterior median with bars showing the 95% posterior intervals. No compositions are shown for years when the spring fishery was closed (1989, 1993–1996, and after 1998).

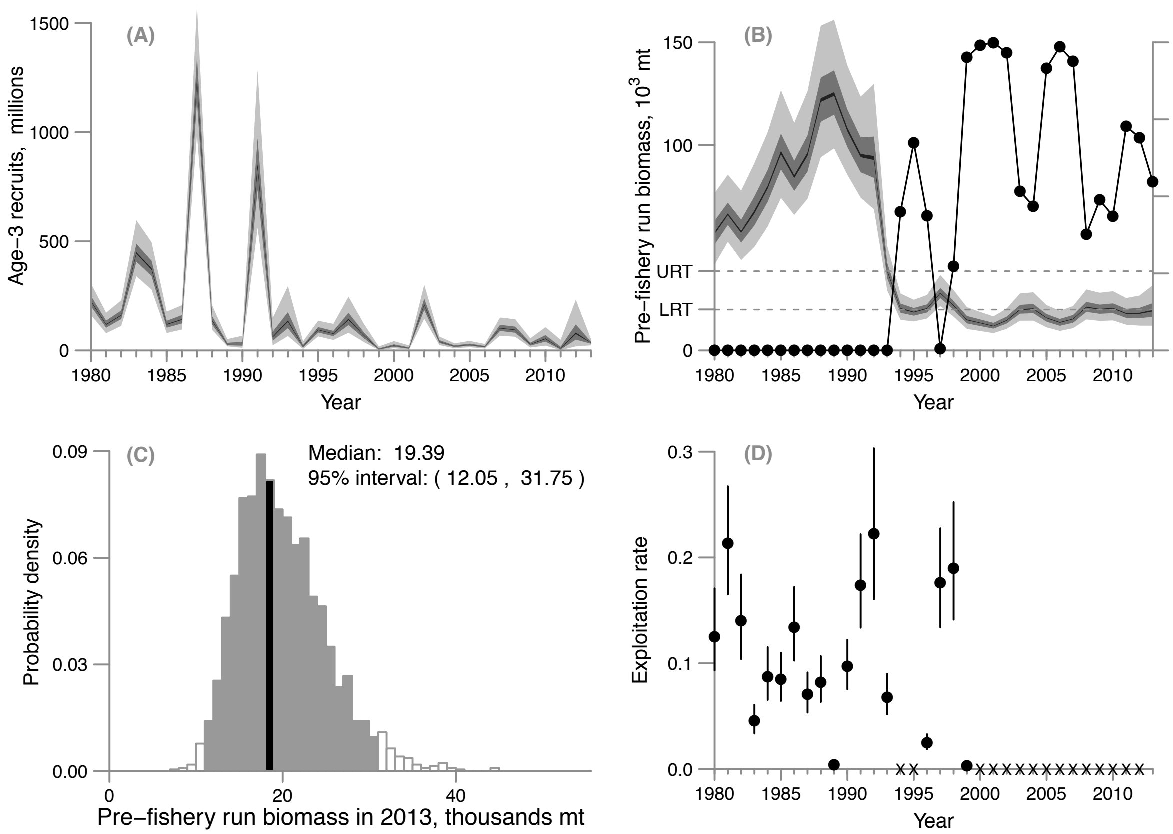

Figure 1.8 The panels show (A) estimated recruitment at age-3 (posterior intervals; light gray = 95% interval, darker gray = 50% interval, black = 5% interval), (B) estimated pre-fishery biomass (posterior intervals; light gray = 95% interval, darker gray = 50% interval, black = 5% interval) and the probability that pre-fishery biomass is below the lower regulatory threshold (LRT) of 22,000 short tons (19,958 mt) (connected black points) with the upper regulatory threshold (URT: 42,500 short tons ≈ 38,555 mt) shown for reference, (C) posterior distribution of estimated pre-fishery biomass for 2013 with the 95% credible interval (light grey) and the median (black) shown, and (D) posterior median exploitation rates (black points) with 95% posterior intervals (segments) – "X" characters represent years the fishery was closed.

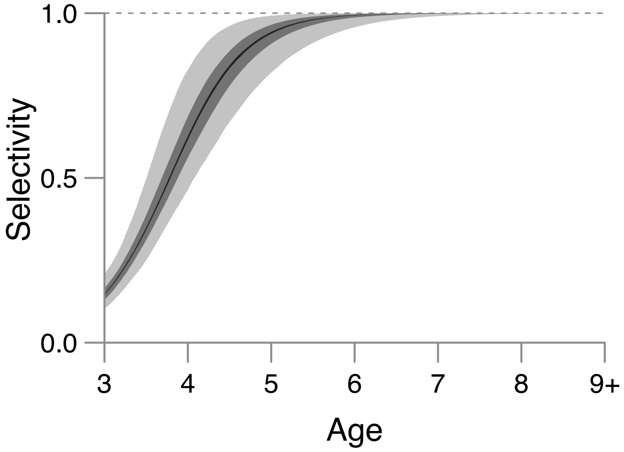

Figure 1.9 Posterior distributions (light gray = 2.5-97.5% interval, darker gray = 25-75% interval, black = median) for model estimates of selectivity proportion at age.

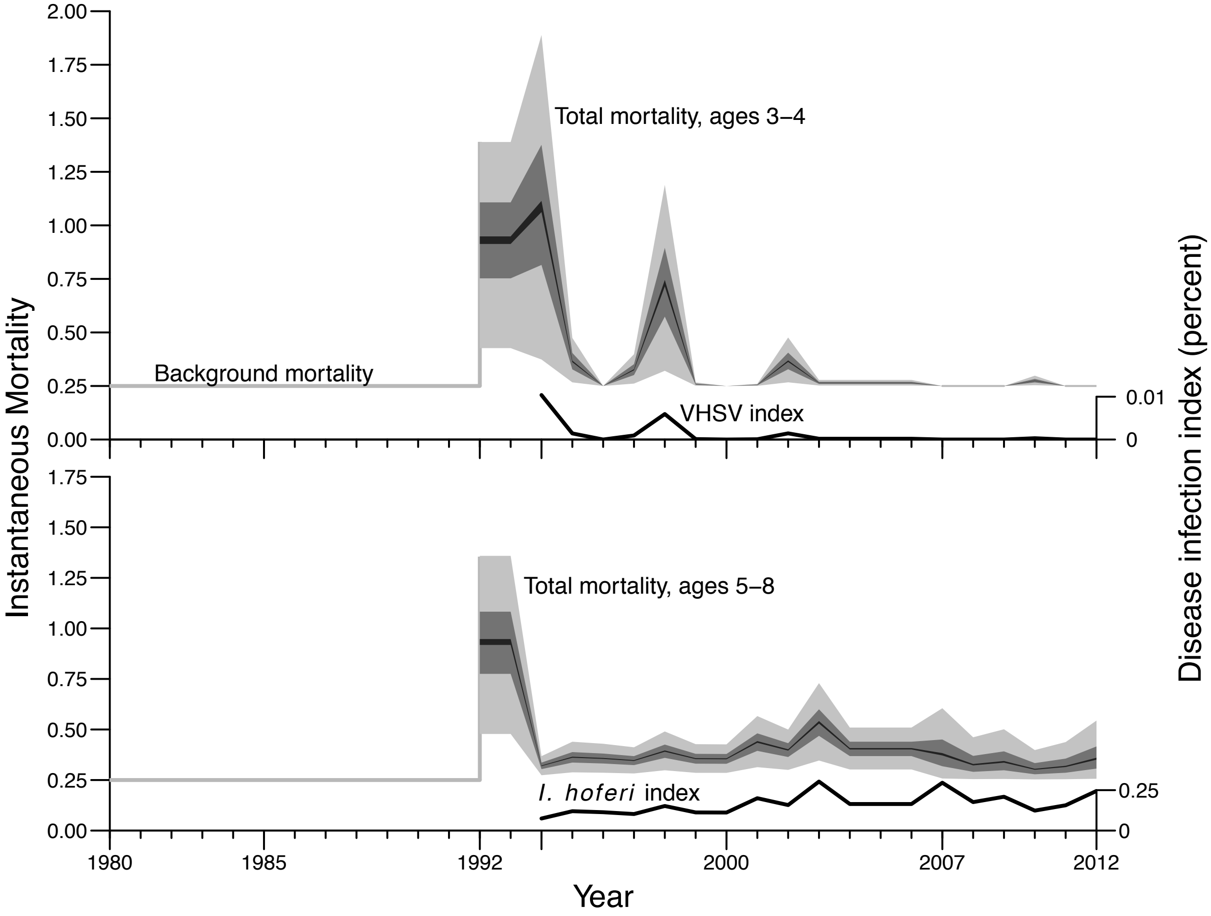

Figure 1.10 Total non-fishery mortality rates for herring of ages 3–8 in 1980–1991 (assumed) and in 1992–2012 (estimated). Posterior distributions begin in 1992 (light gray = 95% interval, darker gray = 50% interval, black = 5% interval) for model estimates of total non-fishery instantaneous mortality for ages 3–4 (top) and ages 5–8 (bottom). Lower black lines show the disease index data for VHSV and *I. hoferi* infection (percent; right-hand axis). VHSV infection was observed to be zero in 2007–2009 and 2011–2012, hence total instaneous mortality in those years equals the assumed background level of 0.25 yr^{-1}.

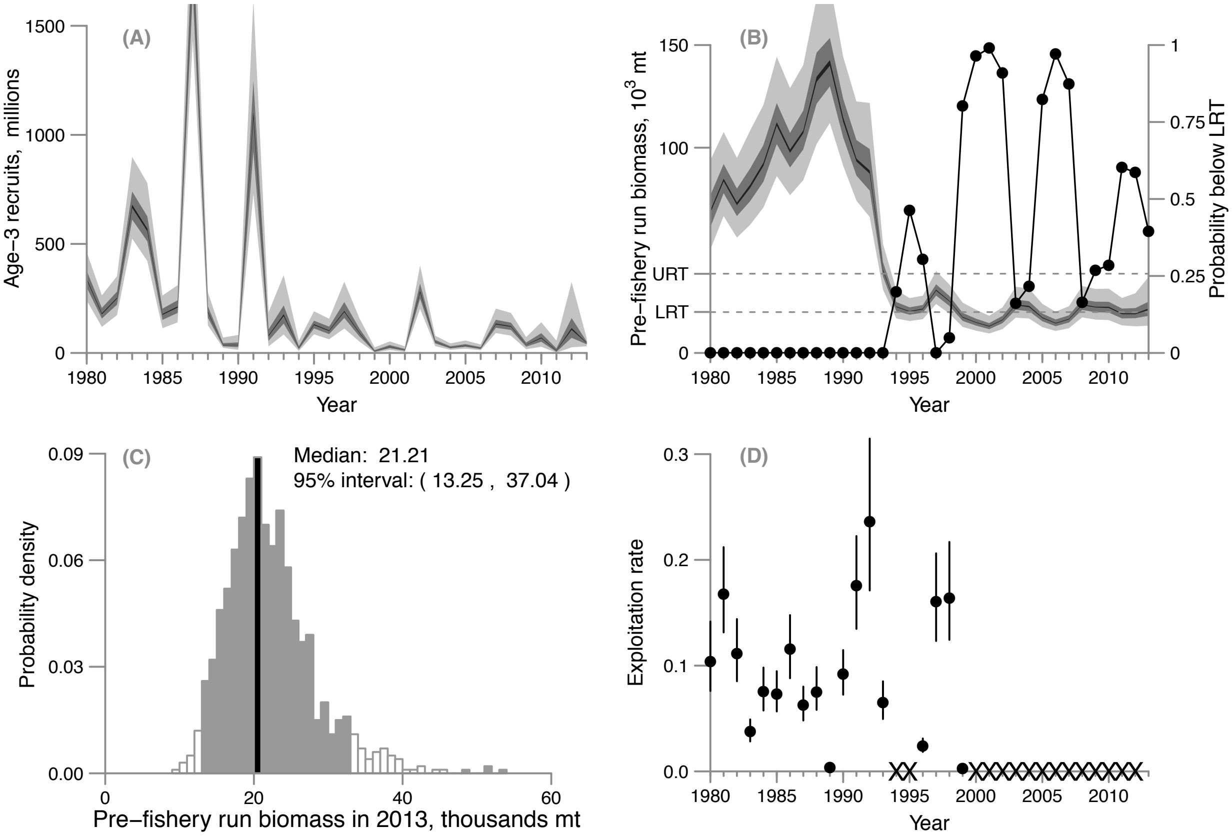

Figure 1.11 Results from the sensitivity analysis using the fixed value of 0.35 yr^{-1} for the background mortality rate.

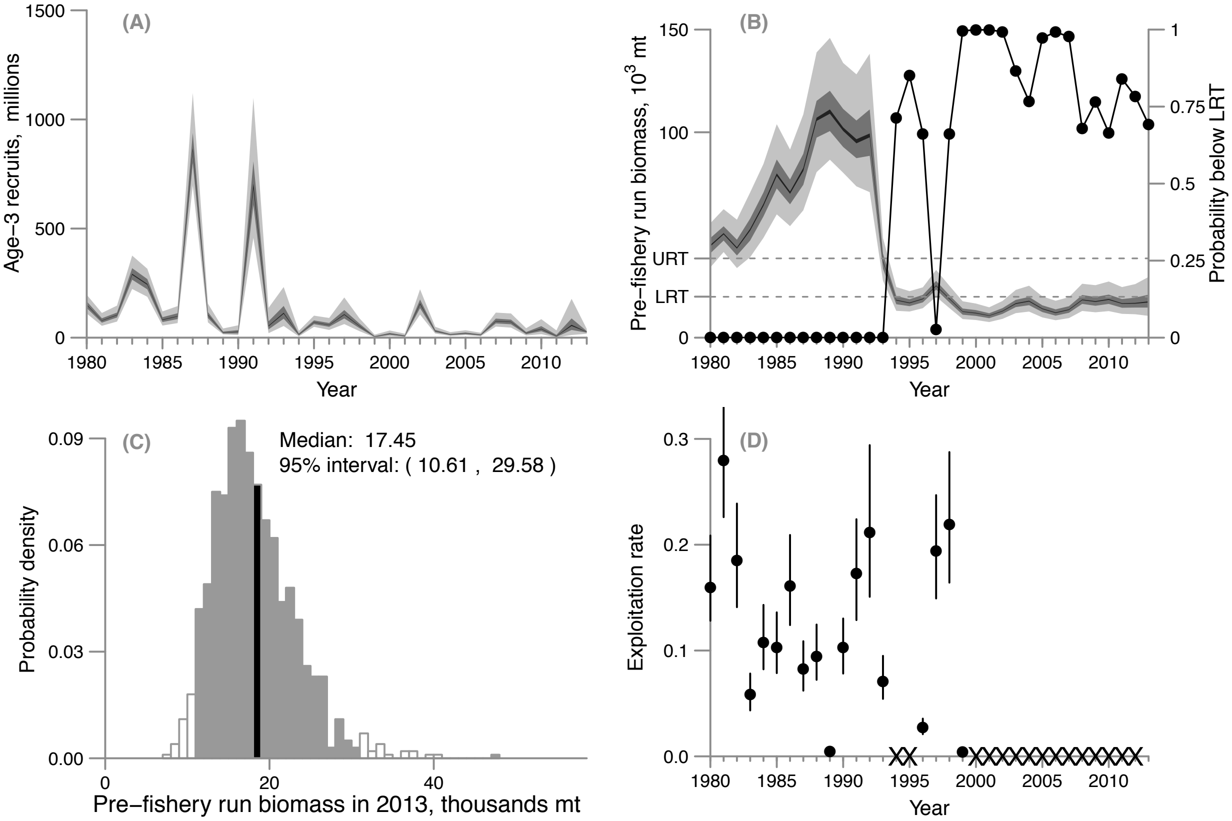

Figure 1.12 Results from the sensitivity analysis using the fixed value of 0.15 yr^{-1} for the background mortality rate.

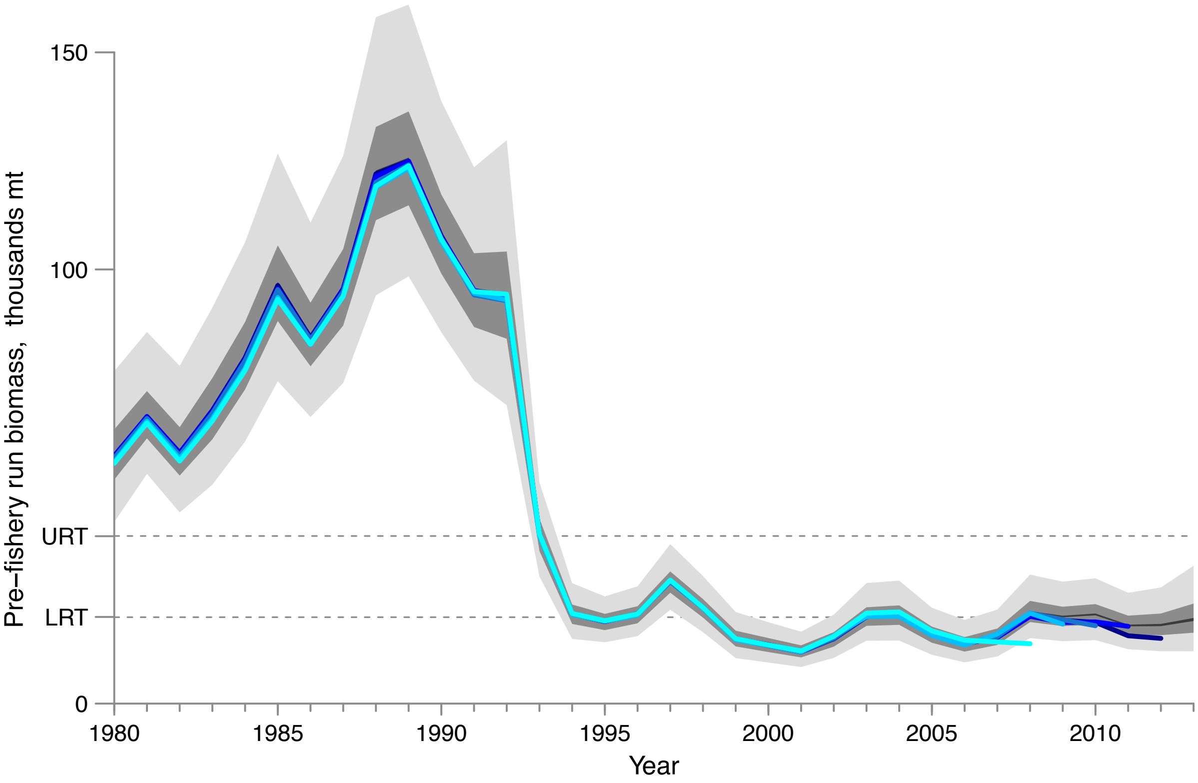

Figure 1.13 Results from five retrospective "peels" compared to the posterior intervals (light gray = 95% interval, darker gray = 50% interval, black = 5% interval) of pre-fishery run biomass from the Bayesian model fit to the entire time series of data. Each "peel" is the posterior median of the model run with an additional year of data removed and lighter blue denotes a peel of data further into the past.

Chapter 2. TRADE-OFFS BETWEEN ACCURACY AND COST OF DATA IN AN INTEGRATED AGE-STRUCTURED ASSESSMENT MODEL

2.1 INTRODUCTION

Stock assessment is the practice of fitting a population dynamics model to data to estimate stock status and regulate harvesting. The foundational theories underlying stock assessment began developing in the early twentieth century. Today, stock assessment is a burgeoning sector of research driven by advancements in sampling technologies, advancements in computing, and a steadily growing human population, which makes more pressing the impetus to sustainably manage our fisheries.

Many managed global fisheries are classified as data limited (having only time series of total landings available to an assessment model, or worse) and numerous other stocks are managed using catch data and at least one index of abundance. However, using multiple data types that include catch information, an index of abundance, and additional time series (such as age or length compositions) can lead to better estimation of stock abundance, productivity, and mortality which are essential for reliable stock assessment predictions (Deriso et al., 1985; Wetzel and Punt, 2011). There are numerous commercially important stocks in the U.S. that have multiple associated scientific surveys, and these collect several data types used to inform more biologically realistic and statistically complex assessment models called integrated assessment models (Fournier and Archibald, 1982; Maunder and Punt, 2013). Additional data types can include female fecundity information, population sex ratio, genetic information, stomach content analysis, predator interaction and abundance, and growth information. It is expensive to conduct the surveys that collect these data types, but little attention is paid to which surveys yield the best return on the money spent in terms of accuracy and precision of management reference points.

The consequences of using inaccurate and imprecise assessment models can be severe; overfishing may lead to the collapse of a population or, on the other hand, underutilization may lead to millions of dollars in economic losses or deficits in food security. It is difficult to determine the accuracy of an assessment model in estimating the true population state in the absence of quantitative analysis. However, simulation studies can be used to fill this gap. Simulation studies involve generating simulations of the true population state, approximating the data sampling process, and then applying the assessment model to each simulation to evaluate the difference between the simulated truth and the estimate from the assessment model.

Numerous studies have investigated the impact of data characteristics on assessment results. For example, Yin and Sampson (2004) tested the impact of different levels of observation error on model output and found that increasing the sample size for age-composition data is most important for accurately estimating ending year biomass. Magnusson and Hilborn (2007) explored the value of trends in catch data and the information contained in age-composition and abundance time series and found that integrated assessment models are surprisingly resilient to a lack of contrast in catch series data, but that age composition data is important for reliably estimating the mortality rate and productivity of the stock. Later, Wetzel and Punt (2011) found that data limited assessments (those using only catch information) showed an improvement in performance after the addition of length-composition data, and Ono et al. (2014) found that the value of data type, quantity, and quality varies by life-history characteristics of the stock, but that an infrequent sampling program run over a longer period is more informative than a shorter, more frequent survey program.

This study builds upon previous work by quantifying the relative value of different data types in terms of return on investment and is motivated by the question of what subset of sampling programs will result in the least loss of accuracy, the lowest estimation bias, and the lowest cost of collection. A general method was developed that can be used to assess the present return on money invested in survey programs to collect different types of data. I conducted a simulation study that employed a leave-one-survey-out approach and looked at the resulting accuracy and precision of model estimates to investigate how data types collected from an individual survey impact management reference points. *A priori*, it was assumed that omitting the most informative data types from an assessment model would lead to the poorest estimation performance.

This method was applied to Pacific herring (*Clupea pallasii*) in Prince William Sound, Alaska, which were historically harvested and commercially valuable, and have been monitored and managed by the Alaska Department of Fish and Game (ADF&G) for over 50 years (Funk and Sandonne, 1990; Thomas and Thorne, 2003). In 1989, the Exxon Valdez oil spill occurred inside of the Sound in late March, during the pre-spawning period of the local herring stock, and herring were adversely affected by the spill that and the following year (Brown and Baker, 1998; Hose et al., 1996; Marty et al., 1999; Norcross et al., 1996). Despite these adverse effects, the population was maintaining relatively high biomass throughout these and subsequent years, but it eventually collapsed in 1992–3 and has yet to recover (Hulson et al., 2008; Pearson et al., 2012; Quinn et al., 2001; Thorne and Thomas, 2008).

Since the spill, and subsequent collapse of the herring population, monitoring of Prince William Sound herring has been a key research objective in the region, resulting in the development of targeted monitoring programs including diver surveys, hydroacoustic surveys, oceanographic data collection, and surveys to assess marine predator abundance and interaction. Thus Prince William Sound herring have a long history of management and have a well-funded monitoring program, making this population an ideal case study to investigate the tradeoffs between data-collection cost and model accuracy. Using the Bayesian age-structured assessment model developed for this population (Chapter 1), I conducted a Monte-Carlo simulation study to explore which data types potentially lead to lower levels of estimation accuracy, consistent bias, and higher probabilities of incorrect management action.

2.2 METHODS

This simulation study used a Bayesian age-structured model as the operating model, which provided simulations of the true state of the population and was used to generate simulations of data with realistic uncertainty. Multiple estimation models (the test scenarios) were fit to these data simulations and model performance was evaluated by comparing estimates to the truth.

2.2.1 *Operating model*

The operating model used in this study was an age-structured population model that simulates herring dynamics based on the parameterization of the Bayesian Prince William Sound herring age-structured assessment model (Table 1.2). The modeling horizon spanned 34 years (1980–2013) and the modeled population included herring of ages 3–8 with a plus group for ages 9 and above. Some proportions of age-3 and age-4 herring were assumed to be mature and cohorts were assumed to be fully mature by age-5. Cohort size was reduced mid-year and at the end of the year by natural mortality (Table index 1.2.1–7) and discreet fishing mortality.

Variation in population process was added to the operating model by drawing sets of parameter values from the posteriors generated by the Markov-chain Monte-Carlo process for the Bayesian Prince William Sound herring assessment model. Using the operating model, each set of parameter values produced a realistic simulation of "true" dynamics in terms of randomly varying recruitment, mortality, and spawning biomass (Fig. 2.1). This step also produced simulations of the "true" indices of biomass and age-compositions from the fishery and survey using the assumed population dynamics (Eqs. 2.1.1–6).

2.2.2 *Data generation*

The following data were assumed to be available to all estimating models without error: catches (from four gear-specific fisheries: gillnet, sac-roe, food and bait, and spawn-on-kelp), empirical weight-at-age, annual sex ratios, and female fecundity-at-age. For the estimation models that assumed disease was impacting the population, disease data (percent infection) were provided without error. Observation error was added to the "true" simulations of biomass indices using bias-corrected, lognormal errors with log standard deviations estimated by the Bayesian assessment model fit to the actual data (Table 1.6 and Table 2.1, Eqs. 2.1.7–12).

Observation error was added to simulated catch-at-age data from the sac-roe fishery and simulated age composition data collected by herring-spawn survey. These data were sampled from the "true" operating model age compositions using a Dirichlet distribution. A Dirichlet distribution was used so that this process accounted for overdispersion in the observed age-composition data due to non-random age- or size-specific spatial variation in pre-spawning and spawning herring aggregations (Hulson et al., 2011). An overdispersion sample size multiplier of 5 was used in the Dirichlet distribution, which brought the effective sample sizes for multinomial-distributed age-compositions to 40 for the sac-roe fishery and 120 for the herring-spawn survey (Eqs. 2.1.9–10). These fishery and survey effective sample sizes were supplied to the EMs.

2.2.3 *Estimation models*

Five estimation models were created, in addition to the base estimation model to examine the relative value of information provided by each survey to the age-structured assessment model.

Comparing worse model performance of a case lacking survey *x* to the base case (which contains survey *x*) reveals the ways in which model performance is improved by the addition of survey *x*.

Base case: this was the reference model that used all collected data types and exactly matches the structure of the operating model (Table 2.2). This case integrated three relative time series of biomass, one absolute index of biomass, and two time series of inter-annual variation in age-specific disease mortality.

Omit disease survey: this model examined the influence of removing the two time series of disease infection used as indices of disease mortality: the protozoan parasite *Ichthyophonus hoferi* and the North American strain of viral hemorrhagic septicemia virus (VHSV). This case reflects the circumstance in which disease-infection data were never collected nor analyzed. In order to design a realistic EM that would result from modeling efforts that never initiated a disease survey, this case assumed that the additional mortality events that contributed to the collapse in 1992–1993 was temporally isolated, and assumes no additional mortality after 1993. This EM also differed from the other cases also by holding plus-group mortality at 1.14 yr^{-1} to address case-specific convergence issues (Table 2.2). This assumption did not impact the relative ranking of this case's EM performance as measured by the metrics listed below (results not shown).

Omit diver survey: tested the value of the diver survey by omitting the egg deposition data (an absolute index of female spawning biomass) and the fecundity-at-age data that were collected in 1984, 1989–93, and 1994–97. A key consequence of this removal is that this EM has no absolute index of biomass, since the egg deposition data are assumed to be equal (in expectation) to female spawning abundance multiplied by fecundity (Table 2.2).

Omit ADF&G hyd. survey: tested the value of the ADF&G hydroacoustic survey, which collected an independent relative index of total biomass of age-3 and older fish each spring from 2005–2009.

Omit PWSSC hyd. survey: tested the value of the PWSSC hydroacoustic survey, which also collected an independent relative index of total biomass of age-3 and older fish each spring, but ran every year from 1995–2012.

Omit aerial survey: tested the value of the aerial survey that ran every year from 1980–2012 by removing the mile-days of milt time series. During the herring spawning events each spring, aerial surveys fly over the spawning sites and measure the linear extent of milt clouds in miles of corresponding coastline per day. Mile-days of milt data were collected every year since 1980 and provide a relative index of male spawning biomass.

The impact of omitting the age/sex/size survey was not investigated since the goal of this study is to identify the value of an individual survey and data from multiple surveys rely on the information collected by the age/sex/size survey. For example, preparation of the hydroacoustics data relies on the length and weight data collected by the age/sex/size survey to convert acoustics data into biomass estimates (Csepp et al., 2011). Therefore, the situation reflected by an EM that removed the age/sex/size survey would defy the premise of this investigation by affecting other survey data as well.

2.2.4 *Survey collection costs*

The values reported (Table 2.3) are the real annual survey costs summed over the years in the time series. Since published budgets include sum of money for programs not directly tied to the cost of running a survey, best estimates of recent annual survey costs used here were obtained by the scientists leading the surveys (Hershberger, P., Moffitt, S., and Pegau, W.S., pers. comm.). Because the time series are patchy, of differing length, and start in different decades adjustment for inflation was necessary to bring past costs to today's currency (2015) so that costs could be compared (Statistics, 2015).

2.2.5 *Convergence diagnostics*

Final results were obtained after 1,758 total model runs: each of the six estimating models was fit to 293 data simulations. Approximately 25% of the data simulations failed to produce a positive definite Hessian, but the individual simulations that failed for one model were different from those that failed for another model, with some overlap. Each estimating model sampled an MCMC chain using the remaining 75% of the data simulations and approximately 30% of those simulations were discarded due to non-convergence. The final 100 data simulations used to produce the final results represent the set of useable simulations that passed convergence diagnostics for all six estimating models simultaneously. Non-convergence was assessed using the Geweke statistic to judge stationarity of the chain, small and decreasing auto-correlation factors for all parameters, and visual assessment of the trace plots. Each of the 600 total retained model runs (for all six estimating models combined) produced a chain of 1,100,000 MCMC iterations with a burn-in of 10%, thinning every thousandth, for a total of 1,000 posterior samples that were used to evaluate performance metrics.

2.2.6 *Reference points*

Three reference points were evaluated as potential management quantities of interest and were chosen for their ability to summarize the assumed dynamics of this population (Table 2.4). The biomass reference points include initial (1980) year biomass and the biomass in 2013 (final-year biomass, the forecast). The two recruitment reference points are mean recruitment for the pre-collapse period (1980-1991) and mean recruitment for the post-collapse period (1992-2013); recruitment in 2013 is a projection. The two mortality reference quantities are mean non-fishery mortality for all ages in the pre-collapse period (1980-1991) and mean non-fishery mortality for all ages in the post-collapse period (1992-2012). There is no projection for non-fishery mortality in 2013 since only pre-fishery run biomass is projected for 2013, which is the biomass of fish at the start of the modeling year before modeled fishery or non-fishery mortality.

2.2.7 *Performance measures*

Model performance was quantified by comparing the estimates of each reference point from the one hundred simulations to the operating model -generated true values using seven performance metrics. These metrics quantify bias, precision, information-to-cost ratios, model reliability, and the probability of false management decisions under each EM.

2.2.7.1 Bias: relative error (RE) and median relative error (MRE)

The median (across MCMC iterations) of the respective joint posterior distribution was used as the estimated value for each reference point ($\hat{\theta}_i$), and the RE of each reference point was calculated as the relative difference between the true value from the operating model (θ_i) and the estimated value ($\hat{\theta}_i$) for each simulation i. The median (MRE) and 90% interval of the distribution of RE values were used to quantify bias and imprecision across simulations when estimating each reference point.

2.2.7.2 Bias and precision: absolute relative error (ARE)

Absolute relative error quantifies the absolute relative difference between the estimated value and the true value from the operating model for each reference point, and is an indicator of both bias and precision. The median (MARE) and 90% interval of the ARE values were used to describe the distribution of ARE values across simulations.

2.2.7.3 Model reliability: extreme tail probability (ETP)

Extreme tail probability detects failures of a Bayesian estimation model in terms of how often the true reference value is within the extreme tails of the posterior distribution (Gelman et al., 2004). For this study, the lower 0.01 and the upper 0.99 percentiles were chosen as the extremes of the posterior distribution. For a well-performing EM it is expected that the true value of each reference point will fall outside of the 98% credible interval in no more than 2% of the simulations. This constraint was relaxed and an ETP rate of greater than 5% was considered an indicator of poor estimation performance since results are based on only 100 data simulations.

2.2.7.4 Probability of false management conclusions (PFM)

The probability of a false closing measures a certain way each EM can fail under a particular management prescription by comparing the posterior median of the forecast from each EM to the management limit and tracking how often the forecast from the EM is below the limit when the true forecast biomass (from the operating model) is above the limit (Eq. 2.5.7). For each EM, the probability of a false closure measures the probability the fishery would be closed due to insufficient estimated biomass when the true biomass level is above the management threshold. False fishery closures are economically important since they may result in unexploited surplus biomass.

The probability of a false opening measures the probability, associated with each EM, that the fishery would be opened due to sufficiently high estimated biomass when the true biomass level is below the management threshold (Eq. 2.5.8). This measures how often the current management practice would lead to false fishery openings under each EM. False fishery openings are ecologically important since they may result in higher exploitation rates than the management strategy prescribes.

The total probability of false management conclusions (PFM) is the number of combined occurrences of false fishery closures and false fishery openings for an EM summed over all data simulations (Eq. 2.5.9).

2.2.7.5 Information-to-cost ratio (ICR)

Each of the EMs differs from the base case in that they each lack a single survey's data (Table 2.2). A metric was developed to quantify the return on money invested in each survey. This metric weighted the improvement in model performance due to each survey's data by that survey's total collection costs (Eq. 2.5.10), and is termed the information-to-cost ratio. Improvement in model performance was quantified by the decrease in MARE when that survey's data were added back to the assessment model. The ICR produces a relative ranking of those surveys that provide a higher return on investment toward estimating accurate and precise management reference points.

2.3 RESULTS

2.3.1 *Which cases influence accuracy and precision of biomass?*

For initial year biomass (B_{1980}), the greatest biases were positive and resulted when the diver survey data were omitted (MRE of 0.18). The next most positive bias occurred when the data from the disease survey were omitted (MRE of 0.16) (Fig. 2.2, panel A). The remaining four cases (including the base case) resulted in low, approximately equal levels of negative bias in B_{1980} (MREs of -0.02 or less). The results for MARE of initial year biomass closely track those of MRE for all cases (Fig. 2.3, panel A).

In general, all cases resulted in more biased (in the same direction) and less precise estimates of B_{2013} than B_{1980} (Figs 2.2 and 2.3, panels A–B). The greatest bias and imprecision in the forecast biomass resulted from omitting the disease survey (MARE of 0.50), followed by omitting the diver survey (MARE of 0.29), and the 90% intervals of ARE for these cases are bounded by 1.29 and 1.10, respectively. The remaining cases (including the base case) estimated B_{2013} with levels of MARE ≤ 0.10 and 90% ARE intervals ≤ 0.60 (Fig. 2.3, panel B).

2.3.2 *Which cases influence accuracy and precision of recruitment?*

The case omitting the diver survey resulted in the greatest bias and imprecision for both recruitment reference points: MARE and 90% intervals of 0.38 (0, 0.86) for R_1 and 0.3 (0, 0.93) for R_2, and MRE for this case reveals positive bias (Fig. 2.3, panels C–D). The remaining cases resulted in similarly low levels of bias and imprecision for R_1 and for R_2 (MARE and 90% intervals $\leq$ 0.17 (0, 0.40)) with the latter reference point estimated with less accuracy and precision than the former; the results for RE and MRE of the recruitment reference points closely track those of MARE for these cases (Fig. 2.2, panels C–D).

2.3.3 *Which cases influence accuracy and precision of non-fishery mortality?*

When estimating mortality in the pre-collapse period ($\overline{M_1}$), all cases had similarly low values of bias and imprecision (all MARE < 0.03 and 90% ARE intervals $\leq$ 0.10) and the results for RE closely track those of ARE (Figs 2.2 and 2.3, panel E). Since non-fishery mortality in the pre-collapse period for all age classes is fixed for the case omitting the disease survey (see Table 2.2), results from this case merely reflect the variation across simulations and should not be compared to the other case results; thus these results are not shown.

The cases omitting the diver surveys resulted in the worst performance, and omitting the disease survey resulted in the next worst performance, for estimating mean mortality in the second period ($\overline{M_2}$) with MARE values of 0.08 and 0.06, respectively. However, the former cases resulted in only a 0.01–0.02 increase in average bias and precision than the remaining four cases.

2.3.4 *Which cases lead to high probabilities of model failure (ETP)?*

The case omitting the disease survey resulted in the highest ETP for all reference points except $\overline{M_1}$ (not estimated; see description of this case) and R_1 (Table 2.6). This case resulted in 56% and 43% probability of model failure when estimating $\overline{M_2}$ and B_{2013}, respectively. Omission of the diver survey resulted in the highest probability of model failure when estimating $\overline{M_2}$ and $\overline{R_1}$ with probabilities 36% and 11%, respectively. For all cases (including the base), $\overline{M_2}$ was estimated least reliably out of all reference points, with ETP values ranging from 8–56%.

2.3.5 *Which cases lead to high probabilities of false management conclusions (PFM)?*

Omitting the disease and the diver surveys resulted in the greatest total probability of false management conclusions (49% and 41%, respectively) and the remaining cases returned total probabilities between 18–25% (Table 2.7, final row). Omitting the disease survey resulted in a 77% probability of a false fishery opening and omitting the diver survey lead to a 58% such probability; both cases resulted in considerably higher probabilities of false openings, rather than false closures. The remaining cases resulted in probabilities of false closures of $\geq$ 24% and probabilities of false openings of $\leq$ 23%, meaning each of these cases had higher probabilities of leading to a false fishery closure, rather than a false fishery opening.

2.3.6 *Which surveys provide the highest return on estimating the forecast quantity?*

The disease survey resulted in the highest return on investment (0.22 ICR in thousands) when estimating B_{2013} and the aerial survey resulted in the next best return on investment, but was 100 times lower than the disease survey's return (Fig. 2.4 top panel). The PWSSC hydroacoustic survey had the relatively lowest return on investment, with an ICR 6 times lower than the aerial survey's ICR.

2.3.7 *Overall results*

All diagnostics pointed to the case omitting the disease survey as the worst performing model and the case omitting the diver survey not performing much better. Overall, these results suggest that the best value for information was contained in the disease and the aerial surveys, which was highly influenced by how cheap these surveys were to run, combined with the fact that the disease and the diver surveys were the most important for assessment model reliability and for avoiding false management conclusions.

2.4 DISCUSSION

Removing the diver survey resulted in the greatest positive bias across biomass, mean recruitment, and mean mortality. The diver survey took place between 1984–97 (Fig. 1.3) and, in the base case, is used as an absolute index of biomass along side three relative biomass indices. When this absolute index was removed, the model overestimated biomass compared to the scale provided by the remaining indices since the suite of data used by the herring model lacks adequate information from the catch curves to estimate mortality. Using the information from the data spanning 1980–1992, the herring biomass appears to have increased while total catches were increasing, which could be explained by a large biomass in the stock. Then once the biomass was low (after 1992), the fishery was closed and the catch time series more or less ended. Therefore, the model lacks adequate contrast in the catch data to estimate the scale of the biomass so that when high biomass levels are proposed during the MCMC iterations, the model can scale recruitment up and estimate q's less than 1 for the remaining time series to fit the trends in the indices. These results reveal that providing the original assessment model with an absolute index of abundance is necessary to keep the scale of the biomass close to the scale of the data.

Removing the information contained in the disease survey led to underestimation in mean recruitment and mean mortality in the post-collapse period ($\overline{M_2}$) and to over estimation of the forecast biomass (Fig. 2.2). Underestimation of $\overline{M_2}$ was due to the assumptions made to develop the EM for this case (Table 2.2): 1992–93 were the only years that involved instantaneous mortality rates greater than the assumed background rate for ages 3–8. As discussed in the Methods section, this case was testing an assessment model that lacked disease data and therefore did not integrate the assumption of increased mortality in the years following the collapse. Thus, the model lacking the disease survey information estimated additional mortality in 1992–93 at a high enough rate to reduce the biomass to fit the sudden decrease in the observed biomass indices. However, without additional mortality in the remaining years, the model was forced to estimate recruitment levels low enough to fit the sustained low biomass levels, leading to negative bias in $\overline{R_2}$. The positive bias in the forecast biomass for this case is explained by the passage of ten years since the large mortality events in 1992–93 combined with low mortality; the population accrued three generations of herring with very low mortality rates and this was sufficient time for the biomass to begin to exceed that of the operating model.

Removal of either hydroacoustic surveys or the aerial survey data resulted in approximately the same relatively moderate increase in bias and imprecision across all management reference points, including the forecast biomass, even though the duration of the three surveys varied greatly. All three surveys display trends that are in agreement so that when one is removed, the other two provide enough information on scale and total biomass for the model to adequately estimate the reference points, on average.

2.4.1 *Which survey data are most important for estimating accurate and precise forecasts?*

All cases resulted in more biased and less precise estimates of B_{2013} compared to B_{1980} due to incomplete cohorts in the age composition data making it difficult to estimate recruitment and mortality in the later years. That being said, the diver and the disease survey should be retained to minimize ARE across all reference points, but especially in the forecast biomass (Table 2.8 top row, and Fig. 2.3). Using results from ETP, omitting the disease survey presented the only significant challenge for the estimating reliable posteriors of the forecast biomass (Table 2.6).

2.4.2 *Omitting which survey data could lead to false fishery openings?*

This question is vital for managers to ensure that their models do not consistently support setting a higher harvest rate than is advised by the management rules. This situation should be avoided to guard against potentially driving a healthy population toward collapse or depleting the biomass below B_{MSY}. Removing the disease survey or the diver survey from the Prince William Sound herring assessment model led to the highest probabilities of false fishery openings when using the posterior median of the forecast biomass to set the harvest rate; therefore retaining these surveys in the assessment model would be advised under conservative management guidelines.

2.4.3 *Omitting which survey data could lead to false fishery closures?*

This question is important for managers to ensure their assessment models do not consistently limit the amount of commercial fishing, resulting in lost revenue and impairing stakeholder relations. Removing the PWSSC hydroacoustic survey or the aerial survey resulted in the highest probabilities of a false fishery closure, therefore retaining the PWSSC hydroacoustic and the aerial surveys would be economically important.

2.4.4 *What are the tradeoffs between cost of data and model accuracy?*

Comparing the separate rankings of information, cost, and ICR shows tradeoffs between cost of data and accuracy in estimation of management reference points (Table 2.8). Simply using a ranking of how expensive it was to run each survey in the most recent year, the disease survey and the aerial survey were least expensive and the ADF&G hydroacoustic survey and the diver survey were the most expensive (Table 2.3, final column). Therefore, an analysis that only considered cost would support continuing the disease and the aerial survey and would support discontinuing either the ADF&G hydroacoustic survey or the diver survey. However, the diver survey was one of the two most informative surveys when considering accuracy and precision of the forecast biomass, and the aerial survey was among the least informative. On the other hand, the aerial survey was important for ensuring against forecasts that lead to false fishery closures, meaning the aerial survey provides return on commercial economic interests outside of the agency.

When considering the ICR, the disease survey unequivocally provided the highest return on investment with the aerial survey providing the next highest return on investment (Figure 3-4, top panel). Even though the diver survey was highly informative for estimating accurate and precise reference points – even more informative than the disease survey for estimating mortality and recruitment – the associated costs were high enough that the resulting return on investment was among the lowest returns.

2.4.5 *Further exploring the tradeoffs between removing the disease survey or the diver survey.*

Removing either the diver survey or the disease survey results in similar total probabilities of drawing a false management conclusion (Table 2.7 bottom row), but the cost savings associated with removing the diver survey are much greater compared to the disease survey. Therefore removing the diver survey results in similar probabilities of false management conclusions, but results in much greater cost savings than removing the disease survey. Regarding estimation accuracy of reference points, removing the disease survey would result in the least accurate forecast biomass, but removal of the diver survey results in the least accurate recruitment and mortality estimates.

When considering precision and accuracy per dollar invested (ICR), the disease survey is the most valuable and the diver, PWSSC hydroacoustic, and ADF&G hydroacoustic surveys are the least valuable. However, an EM without the diver survey would lead to supporting a false fishery opening with greater than 50% probability. Therefore, using only the results from the ICR would support a management decision that places short-term economic gain above cautionary and conservative management. In addition, eliminating the diver survey would have, on average, resulted in 25% positive bias in the forecast biomass and could lead to posterior medians of the forecast biomass that are twice as large as the true biomass levels.

2.4.6 *General methodology and future directions*

The proceeding discussion underscores the importance of assessing the value of data collection programs individually and that of using multiple criteria in such an assessment. The strength of this study's design is the ability to ascertain the value of a single survey at a time; omitting all of the relative indices of abundance at one time would have led to degradation in the results, and the inability to detect which one of the indices was off track compared to the others. Furthermore, assessing the value of each survey using several criteria allows managers to perform multi-objective comparisons similar to the summary of results in Table 2.8. Managers of other stocks and species can adopt heir own set of pertinent performance metrics into the methods outlined in this study to perform a tailored evaluation of their sampling programs.

While there are analytical advantages to omitting a single survey at a time (those discussed above), an obvious extension of this method would be to omit more than one survey at a time. Under this extension, caution should be used when interpreting results since it may be difficult to identify the cause of the resulting estimation performance. Furthermore, investigating the impact of removing more than one survey at a time could become an unwieldy exercise by testing the entire set of subsets of collection programs. Pre-determining subsets of survey data using practical or reasonable criteria would be advised.

The results of this simulation study are based on interpreting the past usefulness of an entire survey as evidence of future utility. These results make use of the information contained in the entire time series collected by a survey and use this as an indication of the future value of the data collected by that survey. While it would be possible to utilize the leave-one-out approach and simulate population dynamics into the future to project the future value of conducting each survey, the results from such an approach would rely on the ability of the study designer to predict and model the management loop and the resulting impact on the population. Simulating the future value of survey data in this way would have been suboptimal for the population used in this study since it would be difficult to predict management reactions to potential changes in stock status going forward, but such an approach may be appropriate for a more stable population whose biomass is sufficiently above the management limit to minimize scenarios.

A third extension of this method would be to test the value of the frequency with which a survey is conducted by creating scenarios that omit years or periods of data from an individual survey more frequently and comparing estimation performance (Ono et al., 2014). This extension of the current study design would give managers additional options for scheduling effort in ways that maintain model performance.

Table 2.1 Key estimated values used in the operating model, with equations, and description of how observation error was added in the data generation step. N() denotes normally-distributed random variables, D() denotes Dirichlet-distributed random variables and c denotes the log-normal correction term $\frac{\boldsymbol{\sigma}^2}{\mathbf{2}}$.

Estimated value	**Estimate and likelihood expression used in EM**	**Eq.**	**Observation error**	**Eq.**
Estimated ADF&G acoustic biomass, mt	$\hat{H}_{1,y} = B_y e^{q_1}$	2.1.1	$\hat{H}_{1,y} = B_y e^{q_1+\varepsilon_{H_1}-c}$ $\varepsilon_{H_1} \sim \mathbf{N}(0, \sigma_{H_1} = 0.29)$	2.1.7
Estimated PWSSC acoustic biomass, mt	$\hat{H}_{2,y} = B_y e^{q_2}$	2.1.2	$\hat{H}_{2,y} = B_y e^{q_2+\varepsilon_{H_2}-c}$ $\varepsilon_{H_2} \sim \mathbf{N}(0, \sigma_{H_2} = 0.35)$	2.1.8
Estimated sac-roe age composition	$\hat{\Theta}_{1,y,a} = \frac{V_a N_{y,a}}{\sum_{a \in A}(V_a N_{y,a})}$	2.1.3	$\hat{\Theta}_{1,y,a} \sim \mathbf{D}\left[(120-1)\left(\frac{V_a N_{y,a}}{\sum_{a \in A}(V_a N_{y,a})}\right)\right]$	2.1.9
Estimated spawning age composition	$\hat{\Theta}_{Sp,y,a} = \frac{\rho_{M,a} N_{y,a}}{\sum_{a \in A}(\rho_{M,a} N_{y,a})}$	2.1.4	$\hat{\Theta}_{Sp,y,a} \sim \mathbf{D}\left[(40-1)\left(\frac{\rho_{M,a} N_{y,a}}{\sum_{a \in A}(\rho_{M,a} N_{y,a})}\right)\right]$	2.1.10
Estimated naturally spawned eggs, trillions	$\hat{E}_y = 10^{-6} \rho_{f,y} \sum_{a \in A} (\tilde{N}_{y,a} f_{y,a})$	2.1.5	$\hat{E}_y = 10^{-6} \rho_{f,y} \sum_{a \in A} (\tilde{N}_{y,a} f_{y,a}) e^{\varepsilon_E - c}$ $\varepsilon_E \sim \mathbf{N}(0, \sigma_E = 0.35)$	2.1.11
Estimated milt, mile-days	$\hat{T}_y = \frac{(1-\rho_{f,y})\tilde{B}_{post,y}}{e^{q_T}}$	2.1.6	$\hat{T}_y = \frac{(1-\rho_{f,y})\tilde{B}_{post,y}}{e^{q_T}} e^{\varepsilon_T - c}$ $\varepsilon_T \sim \mathbf{N}(0, \sigma_T = 0.32)$	2.1.12

Table 2.2 Description of each test estimation model by difference in estimated parameters and constant values; labels for the last five columns refer to the survey being tested by (and therefore omitted from) that case. *Parameter 2.2.6 is defined differently for the model in the case omitting the disease survey; a single additional mortality is estimated for ages 3-8 in 1993 and not assumed to be explicitly associated with disease.

Index	**Parameter**	**Base**	**Disease**	**Diver**	**ADF&G Hyd.**	**PWSSC Hyd.**	**Aerial**
2.2.1	Background mortality, ages 3–8	Fixed at 0.25					
2.2.2	Non-fishery mortality, ages 9^+	$U \sim (0.3, 2.0)$	**Fixed at 1.14**				
2.2.3	VHSV disease scalar, ages 3-4	$U \sim (0.0, 1000)$	**NA**				
2.2.4	*I. hoferi* scalar, ages 5-8, 1994–2006	$U \sim (0.0, 25.0)$	**NA**				
2.2.5	*I. hoferi* scalar, ages 5-8, 2007–12	$U \sim (0.0, 25.0)$	**NA**				
2.2.6	Disease mortality in 1993, VHSV	$U \sim (0.0, 5.0)$	*****				
2.2.7	Disease mortality in 1993, *I. hoferi*	$U \sim (0.0, 5.0)$	**NA**				
2.2.8	Purse-seine gear selectivity	$U \sim (3.0, 5.0)$					
2.2.9	Purse-seine gear selectivity	$U \sim (1.0, 7.0)$					
2.2.10	ADF&G acoustic scalar, log-link	$U \sim (-5.0, 5.0)$			**NA**		
2.2.11	ADF&G acoustic biomass CV	$U \sim (0.0, 0.6)$			**NA**		
2.2.12	PWSSC acoustic scalar, log-link	$U \sim (-5.0, 5.0)$				**NA**	
2.2.13	PWSSC acoustic biomass add'l error	$U \sim (0.0, 0.6)$				**NA**	
2.2.14	Egg deposition additional error	Fixed at 0.3		**NA**			
2.2.15	Milt scalar, log-link	$U \sim (2.3, 7.0)$					**NA**
2.2.16	Milt CV	$U \sim (0.0, 0.6)$					**NA**
2.2.17	Proportion mature at age 3, 1980–96	$U \sim (0.0, 0.75)$					
2.2.18	Proportion mature at age 4, 1980–96	$U \sim (0.0, 1.0)$					
2.2.19	Proportion mature at age 3, 1997–2012	$U \sim (0.0, 1.0)$					
2.2.20	Proportion mature at age 4, 1997–2012	Fixed at 0.9					
2.2.21	Recruitment by year (millions), log-link	$U \sim (0.0, 8.0)$					
2.2.22	Age-4 abundance in 1980, log-link	$U \sim (0.0, 8.0)$					
2.2.23	Age-5 abundance in 1980, log-link	$U \sim (0.0, 8.0)$					

Table 2.3 Information on the survey program being tested by each case: survey being omitted by that case, the data types collected by that survey, the number of years that survey was run (number of years in each time series), the total cost of running the tested survey, and the annual cost of running the most recent survey in the series are shown. Total survey cost was calculated as the sum of the real annual cost of running the survey in each year of the time series with adjustment for inflation.

Survey tested	**Data types collected by survey**	**n**	**Total survey cost, ψ**	**Annual survey cost**
Disease	Index of VHSV/ulcer infection, Index of *I. hoferi* infection	19	\$152,898	\$10,000
Diver	Eggs deposited (trillions) Fecundity (eggs per female)	10	\$4,136,315	\$350,000
ADF&G acoustics	Hydroacoustic biomass (mt)	5	\$292,463	\$67,000
PWSSC acoustics	Hydroacoustic biomass (mt)	20	\$1,354,682	\$80,000
Aerial	Milt (mile-days)	33	\$627,980	\$16,600

Table 2.4 The six reference points used in this study along with their mathematical equation.

Eq.	**Reference point**	**Equation**
2.4.1	Initial year biomass in 1980	B_{1980}, refer to Eq. 1.4.12
2.4.2	Final year biomass in 2013, the forecast	B_{2013}, refer to Eq. 1.4.14–16
2.4.3	Mean recruitment of the pre-collapse period (1980-1991)	$\overline{R_1} = \frac{\sum_{y=1980}^{1991} R_y}{12}$
2.4.4	Mean recruitment of the post-collapse period (1992-2013)	$\overline{R_2} = \frac{\sum_{y=1992}^{2013} R_y}{22}$
2.4.5	Mean mortality of the pre-collapse period (1980-1991)	$\overline{M_1} = \frac{\sum_y \left[\frac{\sum_{a=3}^{9^+} m_a}{7}\right]}{12}$
2.4.6	Mean mortality of the post-collapse period (1992-2012)	$\overline{M_2} = \frac{\sum_y \left[\frac{\sum_{a=3}^{8}\left(m_a + m_{d,a}\right) + m_{9^+}}{7}\right]}{21}$

Table 2.5 Performance measures using the true reference quantity $\boldsymbol{\theta}$ and the estimate $\hat{\boldsymbol{\theta}}$. Subscript *i* is used to index the simulations and *j* to index the MCMC samples for that simulation. In equations 2.5.6–8 the symbol *L* stands for the lower regulatory threshold or limit used to manage the population and *s* stands for 100, the total number of simulations conducted. In equation 2.5.10, the symbol $\Delta(\mathbf{MARE})_{\boldsymbol{h}}$ stands for the difference between the MARE of the base case and the MARE of the case lacking the data from survey *h* and $\boldsymbol{\psi_h}$ stands for the cost of collecting survey *h* (Table 2.3)*;* each of the non-base cases have an associated ICR.

Eq.	**Performance measure**	**Equation**
2.5.1	Bayesian posterior median	$\hat{\theta}_i = \mathrm{median_j}(\hat{\theta}_{ij})$
2.5.2	Relative error	$\mathrm{RE_i} = \frac{\hat{\theta}_i - \theta_i}{\theta_i}$
2.5.3	Median relative error	$\mathrm{MRE} = \mathrm{median_i}\left(\frac{\hat{\theta}_i - \theta_i}{\theta_i}\right)$
2.5.4	Absolute relative error	$\mathrm{ARE_i} = \left\lvert\frac{\hat{\theta}_i - \theta_i}{\theta_i}\right\rvert$
2.5.5	Median absolute relative error	$\mathrm{MARE} = \mathrm{median_i}\left(\left\lvert\frac{\hat{\theta}_i - \theta_i}{\theta_i}\right\rvert\right)$
2.5.6	Extreme tail probability	$\mathrm{ETP} = \frac{\sum_{i=1}^{63}[P(\hat{\theta}_i > \theta_i) < 0.01 \text{ or } P(\hat{\theta}_i > \theta_i) > 0.99]}{s}$
2.5.7	Probability of a false closure	$\mathrm{PFC} = \frac{\sum_i (\hat{\theta}_i < L \text{ and } \theta_i > L)}{\sum_i (\theta_i > L)}$
2.5.8	Probability of a false opening	$\mathrm{PFO} = \frac{\sum_i (\hat{\theta}_i > L \text{ and } \theta_i < L)}{\sum_i (\theta_i < L)}$
2.5.9	Total probability of a false management conclusion	$\mathrm{PFM} = \frac{\sum_i[(\hat{\theta}_i < L \text{ and } \theta_i > L) + (\hat{\theta}_i > L \text{ and } \theta_i < L)]}{s}$
2.5.10	Information-to-cost ratio	$\mathrm{ICR} = \frac{\Delta(\mathrm{MARE})_h}{\psi_h}$

Table 2.6 Values of extreme tail probability (ETP) by reference point and by EM case. High values of ETP translate to a high probability of model failure and values of ETP less than or equal to 0.02 imply model reliability. Darker fill relates to higher values of ETP. The last row displays the mean ETP across reference points for each case.

Reference point	**Base**	**Disease**	**Diver**	**ADF&G Hyd.**	**PWSSC Hyd.**	**Aerial**
B_{1980}	0.01	0.19	0.03	0.03	0.01	0.03
B_{2013}	0.05	0.43	0.04	0.03	0.05	0.02
$\overline{M_1}$	0.02	NA	0.03	0.01	0.02	0.03
$\overline{M_2}$	0.18	0.56	0.36	0.13	0.15	0.08
$\overline{R_1}$	0.05	0.07	0.11	0.04	0.04	0.06
$\overline{R_2}$	0.01	0.12	0.10	0.01	0.00	0.02
Mean ETP	0.05	**0.27**	**0.11**	0.04	0.05	0.04

Table 2.7 Probability of false fishery closures (PFC), probability of false fishery openings (PFO), and total probability of false management conclusions (PFM) resulting from implementing the EM associated with each case. Darker fill relates to higher probabilities among all three measures.

Metric	Base	Disease	Diver	ADF&G Hyd.	PWSSC Hyd.	Aerial
PFC	0.24	0.03	0.13	0.24	0.37	0.29
PFO	0.15	0.77	0.58	0.15	0.15	0.23
PFM	0.18	0.49	0.41	0.18	0.23	0.25

Table 2.8 The best (**) and next best (*) surveys using five criteria that are important for management. Surveys were ranked by how cheap they are, using collection cost in the most recent survey year (Table 2.3, final column), how informative they are, using Δ(MARE) (Fig. 2.4, x-axis of bottom panel), and by their information-to-cost ratio (ICR) relating to the estimate of the forecast biomass, $\boldsymbol{B}_{\mathbf{2013}}$ (Figs 2.4, top panel). Surveys were also ranked by their importance for model reliability (mean ETP: final row of Table 2.6) and their importance to implying correct management decisions (PFM: final row of Table 2.7).

Metric	Criteria	Disease	Diver	ADF&G Hyd.	PWSSC Hyd.	Aerial
Cost	Cheap	**				*
MARE	Informative	**	*			
ICR	High return	**				*
ETP	Reliability	**	*			
PFM	Correct mgmt.	**	*			

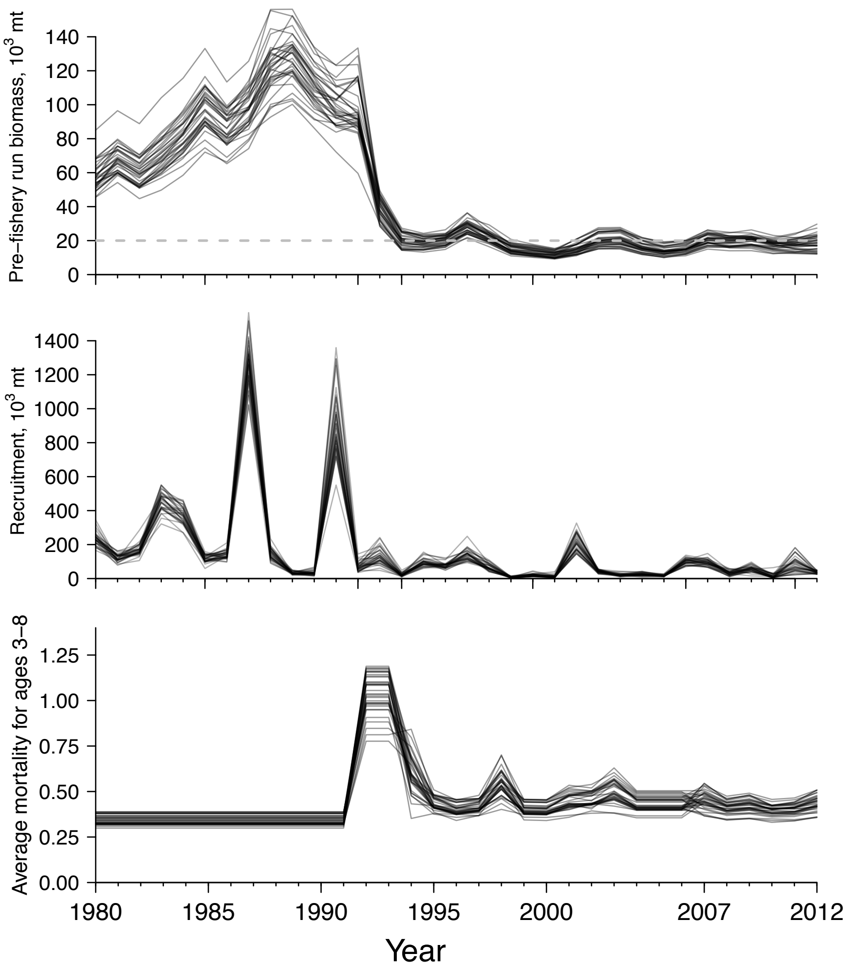

Figure 2.1 The 100 simulations of true biomass, recruitment, and mortality from the operating model. After observation error was added to each simulation, these became the 100 data simulations used by this study.

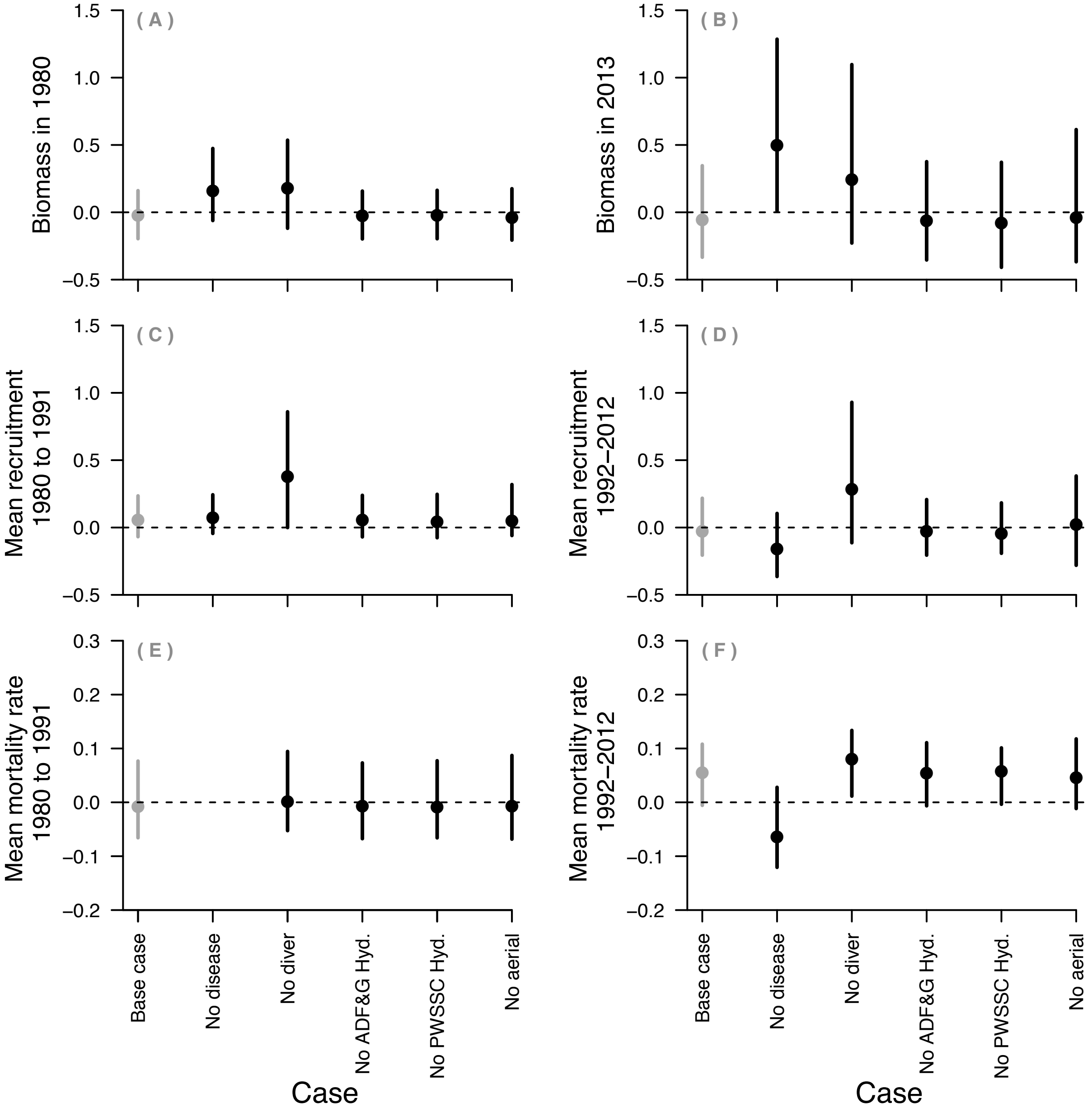

Figure 2.2 Relative error (RE) by case: each panel presents the results across cases by reference point. Filled-in points represent the median and lines illustrate the 90% interval of the RE across simulations. In each panel, the base case is grey.

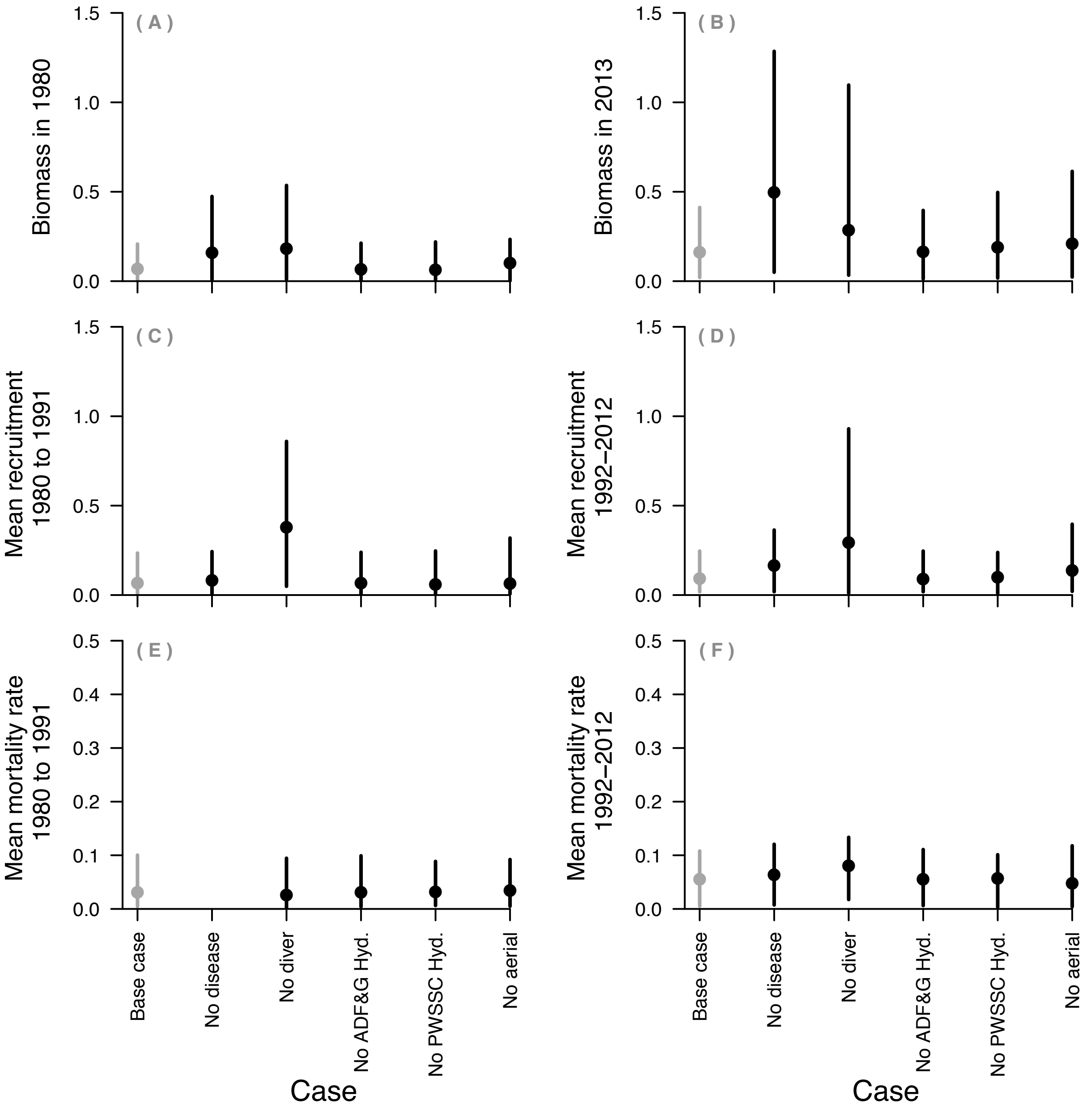

Figure 2.3 Absolute relative error (ARE) by case: each panel presents the results across cases by reference point. Filled-in points represent the median and lines illustrate the 90% interval of the ARE across simulations. In each panel, the base case is grey.

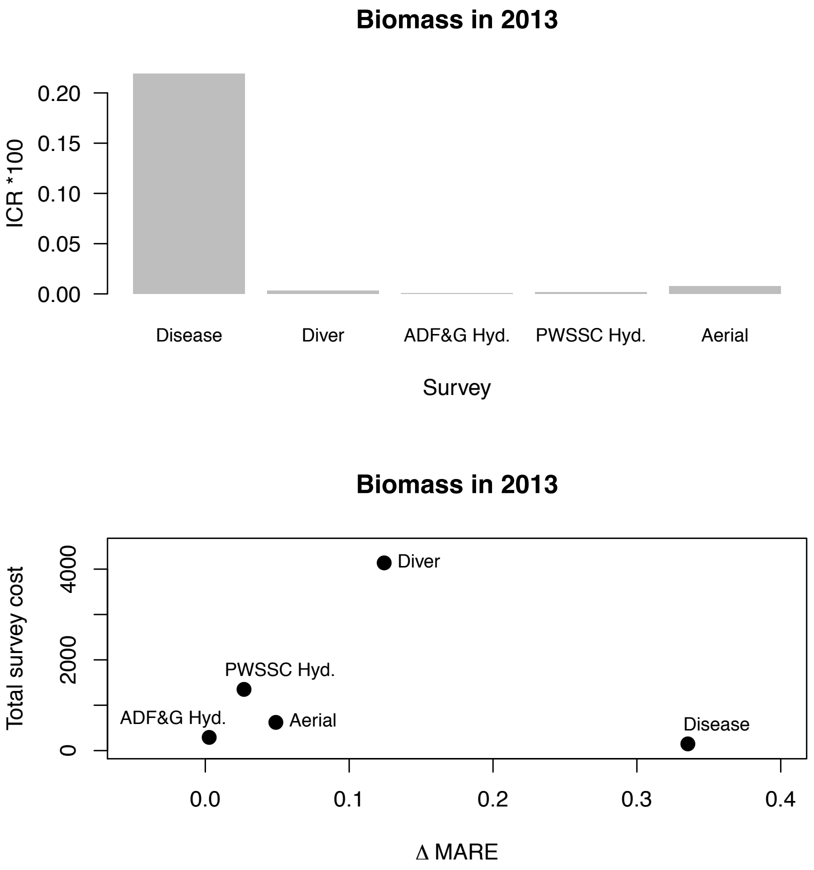

Figure 2.4 The top panel shows the information to cost ratio (ICR) for each survey relating to the estimate of the forecast biomass, $\boldsymbol{B_{2013}}$. The bottom panel shows the cost of each survey program plotted by the improvement in MARE in the forecast biomass due to the addition of that survey's data.

CONCLUSION

The work presented in this thesis developed the first Bayesian assessment of Prince William Sound herring and found that the population remains at low, post-collapse biomass levels and that there is a 54% probability that biomass is below the regulatory limit of 19,958 mt. This work did not seek to determine the driving factors behind this stock's continued low abundance, but instead developed a research tool to aid in these efforts. The Bayesian model produces estimated quantities with measures of uncertainty and probabilities associated with alternative states of the population, which are useful for hypothesis testing. Furthermore, the Bayesian structure for the ASA model allows for the inclusion of further information on this stock as well as information from other herring stocks, in the form of informative, or more informative, priors.

Adoption of the Bayesian model as the agency's assessment tool could also be the basis for managing this population with a decision rule that explicitly considers uncertainty. The amount of risk managers are willing to assume, in terms of strategic economic or conservation goals, can be built directly into a probabilistic regulation rule. Trade-offs with respect to conservation goals and ecosystem balance, as well as economic concerns, and any recommendations should be made after conducting a risk analysis and/or management strategy evaluation using the goals of the agency and the community.

Exploring the tradeoffs between the cost of running each survey and the improvement in model performance due to the inclusion of that survey's data provided a ranking of which historical sampling programs have provided the most valuable data for forecasting biomass. The disease survey (which is relatively cheap and collects an index of additional mortality due to disease) and the diver survey (which is relatively expensive and collects an absolute index of abundance) were found to be the most valuable sampling programs. This work is directly useful to Pacific herring researchers by providing critical information about how to prioritize monitoring efforts going forward. Furthermore, this work presents a general framework for evaluating the return on investment in surveys. Future researchers can use their own set of pertinent performance metrics and adopt the methods outlined in the second chapter to perform a tailored, multi-objective evaluation of their own sampling programs.

BIBLIOGRAPHY

Baker, T. T., and Brady, J. A. 1992. Pacific herring pound spawn-on-kelp fishery in Prince William Sound, Alaska, 1990. Regional Information Report 2A92-02. Ed. by D. o. C. F. Alaska Department of Fish and Game. Anchorage, Alaska.

Barton, L. H., and Wespestad, V. G. 1980. Distribution, biology, and stock assessment of western Alaska's herring stocks. *In* BR Melteff and VG Wespestad (ecis.). Proc. Ak. Herring Symp, pp. 80-84.

Botz, J., Hollowell, G., Bell, J., Brenner, R., and Moffitt, S. 2010. 2009 Prince William Sound area finfish management report. Alaska Department of Fish and Game, Fishery management report No. 10-55, Anchorage, A.K.

Brown, E. D., and Baker, T. T. 1998. Injury to Prince William Sound herring following the *Exxon Valdez* oil spill. *Exxon Valdez* State/Federal Natural Resource Damage Assessment Final Report (Fish/Shellfish Study Number 11). Ed. by D. o. C. F. M. a. D. Alaska Department of Fish and Game. Cordova, Alaska.

Buckland, S. T. 1992. Report of the Scientific Committee, Annex H. Proposal for standard presentation of abundance estimates. Rep. int. Whal. Comm, 42: 235.

Chib, S., and Greenberg, E. 1995. Understanding the metropolis-hastings algorithm. The American Statistician, 49: 327-335.

Csepp, D. J., Vollenweider, J. J., and Sigler, M. F. 2011. Seasonal abundance and distribution of pelagic and demersal fishes in southeastern Alaska. Fisheries Research, 108: 307-320.

Cury, P. M., Boyd, I. L., Bonhommeau, S., Anker-Nilssen, T., Crawford, R. J. M., Furness, R. W., Mills, J. A., et al. 2011. Global seabird response to forage fish depletion—one-third for the birds. Science, 334: 1703-1706.

Deriso, R. B., Quinn II, T. J., and Neal, P. R. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences, 42: 815-824.

Fournier, D. A., and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39: 1195-1207.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., et al. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software, 27: 233-249.

Francis, R. I. C. C. 1992. Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. Canadian Journal of Fisheries and Aquatic Sciences, 49: 922-930.

Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1124-1138.

Funk, F. C., and Sandonne, G. J. 1990. Catch-age analysis of Prince William Sound, Alaska, herring, 1973-1988. Regional Information Report 5J89-02 edn. Ed. by Alaska Department of Fish and Game, Division of Commercial Fisheries. Jouneau, Alaska.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian data analysis*,* Chapman & Hall/CRC, Boca Raton, Florida.

Haegele, C. W., and Schweigert, J. F. 1985. Distribution and characteristics of herring spawning grounds and description of spawning behavior. Canadian Journal of Fisheries and Aquatic Sciences, 42: s39-s55.

Hose, J. E., McGurk, M. D., Marty, G. D., Hinton, D. E., Brown, E. D., and Baker, T. T. 1996. Sublethal effects of the (Exxon Valdez) oil spill on herring embryos and larvae: morphological, cytogenetic, and histopathological assessments, 1989 1991. Canadian Journal of Fisheries and Aquatic Sciences, 53: 2355-2365.

Hulson, P.-J. F., Hanselman, D. H., and Quinn II, T. J. 2011. Determining effective sample size in integrated age-structured assessment models. ICES Journal of Marine Science: Journal du Conseil: fsr189.

Hulson, P.-J. F., Miller, S. E., Quinn, T. J., II, Marty, G. D., Moffitt, S. D., and Funk, F. 2008. Data conflicts in fishery models: incorporating hydroacoustic data into the Prince William Sound Pacific herring assessment model. ICES Journal of Marine Science, 65: 25-43.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., et al. 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science: Journal du Conseil: fsu198.

Kocan, R. M., Marty, G. D., Okihiro, M. S., Brown, E. D., and Baker, T. T. 1996. Reproductive success and histopathology of individual Prince William Sound Pacific herring 3 years after the (Exxon Valdez) oil spill. Canadian Journal of Fisheries and Aquatic Sciences, 53: 2388-2393.

Kopchak, R. J. 2013. The economics of our ecosystem. *In* Delta sound connections. Prince William Sound Science Center, Cordova, AK.

Marty, G. D., Hulson, P.-J. F., Miller, S. E., Quinn, T. J., II, Moffitt, S. D., and Merizon, R. A. 2010. Failure of population recovery in relation to disease in Pacific herring. Diseases of Aquatic Organisms, 90: 1-14.

Marty, G. D., Okihiro, M. S., Brown, E. D., Hanes, D., and Hinton, D. E. 1999. Histopathology of adult Pacific herring in Prince William Sound, Alaska, after the Exxon Valdez oil spill. Canadian Journal of Fisheries and Aquatic Sciences, 56: 419-426.

Marty, G. D., Quinn, T. J., II, Carpenter, G., Meyers, T. R., and Willits, N. H. 2003. Role of disease in abundance of a Pacific herring (*Clupea pallasi*) population. Canadian Journal of Fisheries and Aquatic Sciences, 60: 1258-1265.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61-74.

McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54: 284-300.

McAllister, M. K., Pikitch, E. K., Punt, A. E., and Hilborn, R. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 51: 2673-2687.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science: Journal du Conseil, 56: 473-488.

Monson, D. H., Doak, D. F., Ballachey, B. E., Johnson, A., and Bodkin, J. L. 2000. Long-term impacts of the Exxon Valdez oil spill on sea otters, assessed through age-dependent mortality patterns. Proceedings of the National Academy of Sciences, 97: 6562-6567.

Norcross, B. L., Hose, J. E., Frandsen, M., and Brown, E. D. 1996. Distribution, abundance, morphological condition, and cytogenetic abnormalities of larval herring in Prince

William Sound, Alaska, following the *Exxon Valdez* oil spill. Canadian Journal of Fisheries and Aquatic Sciences, 53: 2376-2387.

Ono, K., Licandeo, R., Muradian, M. L., Cunningham, C. J., Anderson, S. C., Hurtado-Ferro, F., Johnson, K. F., et al. 2014. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science: Journal du Conseil: fsu007.

Parma, A. M. 1993. Retrospective catch-at-age analysis of Pacific halibut: implications on assessment of harvesting policies. *In* Proceedings of the International Symposium on Management Strategies of Exploited Fish Populations. Alaska Sea Grant Rep.

Pearson, W. H., Deriso, R. B., Elston, R. A., Hook, S. E., Parker, K. R., and Anderson, J. W. 2012. Hypotheses concerning the decline and poor recovery of Pacific herring in Prince William Sound, Alaska. Reviews in Fish Biology and Fisheries, 22: 95-135.

Pete, M. 1990. Subsistence use of herring by communities along the eastern Bering Sea coast of Alaska. *In* Proceedings of the International Herring Symposium, pp. 91-01. Anchorage, Alaska.

Peterson, C. H. 2001. The “Exxon Valdez” oil spill in Alaska: acute, indirect and chronic effects on the ecosystem. Advances in Marine Biology, 39: 1-103.

Piatt, J. F., Lensink, C. J., Butler, W., Kendziorek, M., and Nysewander, D. R. 1990. Immediate impact of the Exxon Valdez oil spill on marine birds. The Auk: 387-397.

Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35-63.

Quinn, T. J., II, and Deriso, R. B. 1999. Quantitative fish dynamics*,* Oxford University Press.

Quinn, T. J., Marty, G. D., Wilcock, J., and Willette, M. 2001. Disease and population assessment of Pacific herring in Prince William Sound, Alaska. *In* Herring: expectations for a new millenium. Ed. by F. Funk, J. Blackburn, D. Hay, A. J. Paul, R. Stephensen, R. Toreson, and D. Witherell. University of Alaska Sea Grant, Fairbanks, A.K., Anchorage, A.K.

Reum, J. C. P., Essington, T. E., Greene, C. M., Rice, C. A., Polte, P., and Fresh, K. L. 2013. Biotic and abiotic controls on body size during critical life history stages of a pelagic fish, Pacific herring (*Clupea pallasii*). Fisheries Oceanography, 22: 324-336.

Rooper, C. N., Haldorson, L. J., and Quinn, T. J. 1999. Habitat factors controlling Pacific herring (*Clupea pallasi*) egg loss in Prince William Sound, Alaska. Canadian Journal of Fisheries and Aquatic Sciences, 56: 1133-1142.

Sheridan, T., Botz, J., Wiese, A., Moffitt, S., and Brenner, R. 2013. 2012 Prince William Sound area finfish management report. Ed. by Alaska Department of Fish and Game, Fishery Management Report No. 13-46. Anchorage.

Smith, A. D. M., Brown, C. J., Bulman, C. M., Fulton, E. A., Johnson, P., Kaplan, I. C., Lozano-Montes, H., et al. 2011. Impacts of fishing low–trophic level species on marine ecosystems. Science, 333: 1147-1150.

Smith, A. D. M., Sainsbury, K. J., and Stevens, R. A. 1999. Implementing effective fisheries-management systems – management strategy evaluation and the Australian partnership approach. ICES Journal of Marine Science: Journal du Conseil, 56: 967-979.

Statistics, B. o. L. 2015. National Consumer Price Index, CPI inflation calculator. Bureau of Labor Statistics, U. S. Department of Labor.

Stewart, I. J., Hamel, O. S., and Rose, K. 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. Canadian Journal of Fisheries and Aquatic Sciences, 71: 581-588.

Stokesbury, K. D. E., Foy, R. J., and Norcross, B. L. 1999. Spatial and temporal variability in juvenile Pacific herring, *Clupea pallasi*, growth in Prince William Sound, Alaska. Environmental Biology of Fishes, 56: 409-418.

Teerlink, S. F., von Ziegesar, O., Straley, J. M., Quinn II, T. J., Matkin, C. O., and Saulitis, E. L. 2014. First time series of estimated humpback whale (*Megaptera novaeangliae*) abundance in Prince William Sound. Environmental and Ecological Statistics: 1-24.

Thomas, G. L., and Thorne, R. E. 2001. Night-time predation by Steller sea lions. Nature, 411: 1013-1013.

Thomas, G. L., and Thorne, R. E. 2003. Acoustical-optical assessment of Pacific herring and their predator assemblage in Prince William Sound, Alaska. Aquatic Living Resources, 16: 247-253.

Thorne, R. E., and Thomas, G. L. 2008. Herring and the "Exxon Valdez" oil spill: an investigation into historical data conflicts. ICES Journal of Marine Science, 65: 44-50.

Wetzel, C. R., and Punt, A. E. 2011. Model performance for the determination of appropriate harvest levels in the case of data-poor stocks. Fisheries Research, 110: 342-355.

Yin, Y., and Sampson, D. B. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. North American Journal of Fisheries Management, 24: 865-879.