Special Publication No. 10-01

Evaluation of Stock Assessment and Modeling Options to Assess Sablefish Population Levels and Status in the Northern Southeast Inside (NSEI) Management Area

by

Franz Mueter

January 2010

Alaska Department of Fish and Game **Divisions of Sport Fish and Commercial Fisheries**

Symbols and Abbreviations

The following symbols and abbreviations, and others approved for the Système International d'Unités (SI), are used without definition in the following reports by the Divisions of Sport Fish and of Commercial Fisheries: Fishery Manuscripts, Fishery Data Series Reports, Fishery Management Reports, and Special Publications. All others, including deviations from definitions listed below, are noted in the text at first mention, as well as in the titles or footnotes of tables, and in figure or figure captions.

Weights and measures (metric)

centimeter	cm
deciliter	dL
gram	g
hectare	ha
kilogram	kg
kilometer	km
liter	L
meter	m
milliliter	mL
millimeter	mm

Weights and measures (English)

cubic feet per second	ft^3/s
foot	ft
gallon	gal
inch	in
mile	mi
nautical mile	nmi
ounce	oz
pound	lb
quart	qt
yard	yd

Time and temperature

day	d
degrees Celsius	°C
degrees Fahrenheit	°F
degrees kelvin	K
hour	h
minute	min
second	s

Physics and chemistry

all atomic symbols	
alternating current	AC
ampere	A
calorie	cal
direct current	DC
hertz	Hz
horsepower	hp
hydrogen ion activity (negative log of)	pH
parts per million	ppm
parts per thousand	ppt, ‰
volts	V
watts	W

General

Alaska Administrative Code	AAC
all commonly accepted abbreviations	e.g., Mr., Mrs., AM, PM, etc.
all commonly accepted professional titles	e.g., Dr., Ph.D., R.N., etc.
at	@
compass directions:	
east	E
north	N
south	S
west	W
copyright	©
corporate suffixes:	
Company	Co.
Corporation	Corp.
Incorporated	Inc.
Limited	Ltd.
District of Columbia	D.C.
et alii (and others)	et al.
et cetera (and so forth)	etc.
exempli gratia (for example)	e.g.
Federal Information Code	FIC
id est (that is)	i.e.
latitude or longitude	lat. or long.
monetary symbols (U.S.)	$, ¢
months (tables and figures): first three letters	Jan,...,Dec
registered trademark	®
trademark	™
United States (adjective)	U.S.
United States of America (noun)	USA
U.S.C.	United States Code
U.S. state	use two-letter abbreviations (e.g., AK, WA)

Measures (fisheries)

fork length	FL
mideye to fork	MEF
mideye to tail fork	METF
standard length	SL
total length	TL

Mathematics, statistics

all standard mathematical signs, symbols and abbreviations	
alternate hypothesis	H_A
base of natural logarithm	*e*
catch per unit effort	CPUE
coefficient of variation	CV
common test statistics	(F, t, χ^2, etc.)
confidence interval	CI
correlation coefficient (multiple)	R
correlation coefficient (simple)	r
covariance	cov
degree (angular)	°
degrees of freedom	df
expected value	*E*
greater than	>
greater than or equal to	≥
harvest per unit effort	HPUE
less than	<
less than or equal to	≤
logarithm (natural)	ln
logarithm (base 10)	log
logarithm (specify base)	$\log_2$, etc.
minute (angular)	'
not significant	NS
null hypothesis	H_O
percent	%
probability	P
probability of a type I error (rejection of the null hypothesis when true)	α
probability of a type II error (acceptance of the null hypothesis when false)	β
second (angular)	"
standard deviation	SD
standard error	SE
variance	
population	Var
sample	var

SPECIAL PUBLICATION NO. 10-01

EVALUATION OF STOCK ASSESSMENT AND MODELING OPTIONS TO ASSESS SABLEFISH POPULATION LEVELS AND STATUS IN THE NORTHERN SOUTHEAST INSIDE (NSEI) MANAGEMENT AREA

by

Franz Mueter,
Sigma Plus Statistical Consulting
1085 Arctic Circle
Juneau, AK, 99801

Alaska Department of Fish and Game
Division of Sport Fish, Research and Technical Services
333 Raspberry Road, Anchorage, Alaska, 99518-1599

January 2010

The Special Publication series was established by the Division of Sport Fish in 1991 for the publication of techniques and procedures manuals, informational pamphlets, special subject reports to decision-making bodies, symposia and workshop proceedings, application software documentation, in-house lectures, and became a joint divisional series in 2004 with the Division of Commercial Fisheries. Special Publications are intended for fishery and other technical professionals. Special Publications are available through the Alaska State Library, Alaska Resources Library and Information Services (ARLIS) and on the Internet: http://www.sf.adfg.state.ak.us/statewide/divreports/html/intersearch.cfm. This publication has undergone editorial and peer review.

Franz Mueter
Sigma Plus,
1085 Arctic Circle, Juneau, Alaska 99801, USA

This document should be cited as:
Mueter, F. 2010. Evaluation of stock assessment and modeling options to assess sablefish population levels and status in the Northern Southeast Inside (NSEI) management area. Alaska Department of Fish and Game, Special Publication No. 10-01, Anchorage.

The Alaska Department of Fish and Game (ADF&G) administers all programs and activities free from discrimination based on race, color, national origin, age, sex, religion, marital status, pregnancy, parenthood, or disability. The department administers all programs and activities in compliance with Title VI of the Civil Rights Act of 1964, Section 504 of the Rehabilitation Act of 1973, Title II of the Americans with Disabilities Act (ADA) of 1990, the Age Discrimination Act of 1975, and Title IX of the Education Amendments of 1972.

If you believe you have been discriminated against in any program, activity, or facility please write:

ADF&G ADA Coordinator, P.O. Box 115526, Juneau, AK 99811-5526
U.S. Fish and Wildlife Service, 4401 N. Fairfax Drive, MS 2042, Arlington, VA 22203
Office of Equal Opportunity, U.S. Department of the Interior, 1849 C Street NW MS 5230, Washington DC 20240

The department's ADA Coordinator can be reached via phone at the following numbers:

(VOICE) 907-465-6077, (Statewide Telecommunication Device for the Deaf) 1-800-478-3648,
(Juneau TDD) 907-465-3646, or (FAX) 907-465-6078

For information on alternative formats and questions on this publication, please contact:

ADF&G Division of Sport Fish, Research and Technical Services, 333 Raspberry Road, Anchorage AK 99518 (907) 267-2375.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF APPENDICES

LIST OF APPENDICES (Continued)

LIST OF APPENDICES (Continued)

ABSTRACT

A review of assessment approaches currently used for other Northeast Pacific groundfish stocks suggests that developing an age-structured model for sablefish (*Anoplopoma fimbria*) in the Northern Southeast Inside management area ("Chatham Strait") of southeast Alaska is appropriate and provides a reasonable option for integrating much of the available data. The age-structured model provides a reconstruction of historical abundances, an evaluation of current stock status relative to historical and unfished levels, and estimates of important management parameters. Preliminary results suggest that model results are sensible and that the biomass of sablefish in Chatham Strait is considerably lower than the biomass that would produce maximum sustainable yield. However, the age-structured assessment (ASA) assumes that the Chatham Strait population is "closed" (no immigration/emigration). The validity of this assumption is questionable, as the Chatham Strait population may be part of a single population ranging from British Columbia to the Bering Sea. Therefore, I recommend that a mark-recapture model be used to provide fishery-independent estimates of absolute abundance in Chatham Strait at the time of the fishery. Although tagging data and mark-recapture models suggest that both emigration from and extensive immigration into Chatham Strait occurred throughout late summer and fall of 2006, migration is not accounted for in the current stock assessment approach. Extensive movements of sablefish between Chatham Strait and outside waters, and similarities in population trends suggest that sablefish should be managed as a single coast-wide population. Therefore, I recommend that an ASA for Chatham Strait be developed in collaboration with NOAA Fisheries and DFO Canada, with the long-term goal of developing a single sablefish model that accounts for migration between inside and outside waters. For now, mark-recapture estimates of abundance should be used in conjunction with the ASA. Future research efforts should be directed towards resolving the movement of sablefish into and out of Chatham Strait so it can be adequately accounted for in the assessment.

Key words: Sablefish, *Anoplopoma fimbria*, Chatham Strait, age-structured model, stock assessment, mark-recapture analysis, NSEI

INTRODUCTION

SCOPE OF WORK

A contract was established with Dr. Franz Mueter, Sigma Plus, to evaluate current stock assessment methods for Northern Southeast Inside Subdistrict (NSEI) sablefish (*Anoplopoma fimbria*) and to explore and compare alternative options. The scope of work under the contract included the following:

- Review past and current stock assessment and modeling approaches used for NSEI sablefish.
- Evaluate and compare estimation methods, models and approaches used elsewhere in the North Pacific for sablefish and in other regions worldwide for populations with similar characteristics.
- Provide a comprehensive written document that details the considerations of various prospective models and stock assessment techniques and recommends the most suitable direction to take, based on the contractor's professional opinion.
- Provide a programmed template of the estimator or model identified as the most promising, coded in software agreed upon by the contractor and the Region I Groundfish Biometrician.

In this document, I summarize and compare stock assessment approaches used elsewhere for sablefish or similar species, present several alternatives for an analysis of tagging data from the 2006 tagging program, and present the structure and results from an age-structured analysis of the Chatham Strait sablefish stock for 1980–2006. In addition, I provide a separate spreadsheet and WinBUGS (Spiegelhalter et al. 1995) code for implementing the tagging models, as well as a spreadsheet and an AD Model Builder (ADMB; Otter Research 2007) implementation of the age-structured model. Detailed descriptions of these models and instructions to run the models are included in several appendices.

REVIEW AND EVALUATION OF STOCK ASSESSMENT APPROACHES

OBJECTIVES OF STOCK ASSESSMENT

Stock assessments of sablefish in Chatham Strait have been used to inform management about the status of the stock and set appropriate harvest levels (Carlile et al. 2002). General objectives of a stock assessment are to evaluate current stock status, to evaluate the mean productivity of a stock to determine sustainable harvest levels, and/or to reconstruct the history of the stock (i.e., historical abundances or biomass, recruitment, growth patterns, etc.) (Walters and Martell 2004). I briefly describe these objectives with reference to the current approach used for Chatham Strait sablefish.

Evaluating Current Stock Status

In most managed stocks, the status of a stock is evaluated to estimate the impact of the current level of harvesting on the stock, as measured by the exploitation rate, relative to some goal or standard. A harvest control rule is used to determine the desired catch (C) or exploitation rate (u), which may be constant or may be allowed to vary as a function of abundance or biomass (B). A target exploitation rate may be obtained from a yield-per-recruit analysis, other simulation analyses, or from experience with similar stocks. The actual exploitation rate can be estimated if an estimate of biomass is available ($u = C/B$) or can be determined directly from tagging studies. Direct estimates of exploitation rate were obtained for Chatham Strait sablefish for several years because absolute biomass estimates were deemed unreliable (Carlile et al. 2002).

Currently, the Chatham Strait sablefish is managed using the following approach:

1) Estimate biomass using mark-recapture analysis (modified Petersen estimator)
2) Forecast the number of fish and biomass for the upcoming year, along with the associated lower and upper confidence limits (e.g., B_{lower})
3) Apply $F_{40\%}$ harvest control rule:
    a) Determine u corresponding to the $F_{40\%}$ rule (Yield-per-recruit calculation)
    b) Determine the allowable catch: Catch = $u * B_{\text{lower}}$

The $F_{40\%}$ strategy consists of fishing at a rate that reduces spawning biomass per recruit to 40% of the theoretical, unfished value. The rule is based on stochastic simulations which show that, for groundfish stocks with “typical” life history characteristics, a high average yield can be assured when fishing at the $F_{40\%}$ rate (Clark 1991, 1993). A modified $F_{40\%}$ rule has been used successfully for many groundfish stocks in the US (NPFMC 2007), including sablefish in the Gulf of Alaska (Hanselman et al. 2006). Sablefish off the US West Coast (Washington, Oregon and California) are currently managed using a more conservative $F_{45\%}$ harvest rate (Schirripa and Colbert 2005).

Evaluating Mean Productivity

The mean productivity of a stock, which depends on the long-term relationship between fishing mortality, stock size, and yield, provides a basis for determining sustainable yield. Mean productivity may be estimated from the rate of population growth (e.g., based on a surplus production model) or from the number of recruits at low levels of biomass (based on a spawner-recruitment relationship). If mean productivity is known, estimates can be made of maximum sustainable yield (MSY), the biomass at which MSY is obtained (B_{MSY}), and the fishing mortality ($\mathrm{F}_{\mathrm{MSY}}$) that, on average, achieves MSY and reduces biomass to B_{MSY}. Such a determination is often quite uncertain and has not been done for Chatham Strait, to my knowledge. However, $\mathrm{F}_{40\%}$ provides what is believed to be a conservative proxy for $\mathrm{F}_{\mathrm{MSY}}$. The use of $\mathrm{F}_{40\%}$ in other fisheries is discussed below.

Stock Reconstruction

Stock reconstruction attempts to determine how production (population growth, recruitment) has varied over time and what has caused these changes. Factors affecting production include stock size (density-dependent effects), fishing, and environmental variability. Most stock assessment models currently in use only include the effects of stock size and fishing. However, the stock assessment of sablefish off the US West Coast is one of the few that incorporates the effects of environmental variability on recruitment (Schirripa and Colbert 2006).

Modern stock reconstructions typically include a variety of data types that are often integrated in a complex age-structured model. Such models incorporate data on abundance (from fishery-independent surveys and/or fishery-based indices), growth, mortality, and recruitment, thus including considerable biological realism. A stock reconstruction implicitly evaluates current stock status and may also provide estimates of mean productivity (hence MSY) if, for example, a reliable stock-recruitment relationship is included in an age-structured assessment. Therefore it can address all three of the main objectives of a stock assessment. In spite of the widespread use of complex, age-structured models, which have gained popularity due to the availability of aging data and powerful software to fit the models, their performance in a management context is not necessarily better than that of simpler, less realistic models.

Several stock reconstructions based on an age-structured model have been attempted for the Chatham Strait sablefish stock, most recently in 2002. However, a mark-recapture estimate of abundance was ultimately used to set quotas because of concerns over sablefish movements into and out of Chatham Strait and because of large uncertainty in the estimates of absolute population size (Dressel, personal communication[1]). As discussed below, the same concerns are still valid for an updated age-structured analysis, but these concerns can be somewhat alleviated by “tuning” the ASA to mark-recapture estimates of abundance. Although the results from the ASA were not used because the biomass estimate was not deemed reliable, the decreasing trend in the population biomass that was evident in the ASA seems to have contributed to recent concerns over a declining population.

---

[1] Dressel, S.C., Alaska Department of Fish and Game, Region I, Division of Commercial Fisheries, P.O. Box 110024, Juneau, Alaska 99811-0024.

HARVEST POLICIES

I provide a very brief overview of commonly used harvest policies to provide a context for the current Chatham Strait policy, which is essentially a constant harvest rate policy that applies a harvest rate corresponding to the $F_{40\%}$ rule to the estimated biomass, regardless of the level of biomass. Additional precaution has been applied on an *ad-hoc* basis by applying the desired harvest rate to the lower limit of a 90% confidence interval for sablefish biomass—instead of applying it to the best point estimate.

Commonly used harvest policies in other fisheries include:

Surplus production policy: Under this policy any “excess” biomass that can be harvested without reducing biomass (surplus production) is estimated for a given year and the catch is set at or below this level. A surplus production policy has been used for Pacific halibut in the past. A disadvantage of this policy is that estimates may be highly variable, resulting in a large variability in quotas from year to year.

Constant harvest policy: A determination of a low-risk/optimal long-term sustainable catch is made once (based on simulations or other information) and is applied to all future years. This can be risky if biomass decreases to low levels for any reason.

Constant harvest rate policy: Harvest is set proportional to an estimate of biomass ($\text{Catch} = u * B$), which is similar to the rule currently used for Chatham Strait sablefish, although it does not appear to be formalized or fixed in regulation. Variants of this policy are probably the most commonly applied policies today and have been shown to result in high long-term yields, while being robust to environmental variability (Parma 1990, Parma and Deriso 1990, Walters and Parma 1996).

Threshold policy: No catch is allowed if biomass is below a fixed threshold level, and a constant harvest rate is applied if biomass is above the threshold level. A fixed escapement policy, widely and successfully used for salmon in the State of Alaska, is a special case of a threshold policy with the threshold corresponding to the fixed escapement goal and the harvest rate is set to 1 if abundance exceeds the threshold. This requires good estimates of abundance or biomass (total abundance or escapement) and catches can be highly variable.

Adjustable rate policy: Adjustable rate policies apply rules that make the desired harvest rate dependent on the estimated stock status (biomass). For example, for most groundfish stocks the North Pacific Fisheries Management Council applies an adjustable rate policy that applies a constant harvest rate if the spawning biomass exceeds the biomass that produces maximum yield (B_{MSY} or a proxy like $B_{40\%}$) and reduces the harvest rate linearly if biomass drops below this reference point (Figure 1). The harvest rate is reduced to zero when biomass decreases below a fixed fraction (e.g., 5%) of the unfished biomass. In contrast, the current ADF&G harvest control rule applies the $F_{40\%}$ harvest control rule regardless of the level of absolute biomass.

STOCK ASSESSMENT APPROACHES

I evaluated the following assessment approaches as options for assessing the sablefish population in Chatham Strait:

- Mark-recapture estimation
    - Single-capture event (Petersen/Chapman)
    - Multiple capture events (within/between years)
    - Various other extensions to relax assumptions
    - Time series approach to "smooth" mark-recapture estimates
- Survey estimates (relative index or absolute biomass estimate)
- Depletion estimator
- Surplus production model
- Delay-difference model
- Age-structured assessment model

Each of the approaches is briefly described and advantages and disadvantages for assessing sablefish in Chatham Strait are evaluated. Based on this initial evaluation and on the history of past approaches used in Chatham Strait, I only pursued mark-recapture approaches and an age-structured assessment model in more detail. Hence, these approaches are discussed in detail in separate sections below. First, I briefly describe other commonly used approaches.

Mark-Recapture Estimation

The basic principle of all mark-recapture estimates relies on marking a known number of fish, resampling the population, and assuming that the proportion of marked fish (*M*) in the total population (*N*) is reflected in the proportion of marked fish (*m*) in a sample of fish of size *n* obtained after the marking event. A simple estimator of abundance (the Petersen estimator) can therefore be obtained as follows:

$$\frac{M}{N} = \frac{m}{n} \qquad \Rightarrow \quad \hat{N} = \frac{M \cdot n}{m}$$

A number of restrictive assumptions have to hold for the estimator to be valid, most notably that the population is closed (no mortality, emigration, immigration, recruitment, etc.), that tagged and untagged fish have the same probability of being captured in the resampling event, and that fish do not lose their tags (Seber 1982, Williams et al. 2002). Some of these assumptions can be relaxed by modifying the estimator appropriately as shown in detail in the mark-recapture section.

Mark-recapture estimates of sablefish abundance in Chatham Strait have been obtained from a tagging program that began in 1997. Abundance estimates are converted to biomass based on the age composition and average weight-at-age data and a yield-per-recruit analysis is used to estimate the fishing mortality rate ($F_{40\%}$) that, in the long run, will reduce sablefish spawning biomass to 40% of the unfished spawning biomass.

Survey Estimates

As an alternative to mark-recapture estimates, area-swept estimates using a trawl could provide reliable estimates of abundance if most of Chatham Strait is trawlable and could reasonably be surveyed (or if the proportion of trawlable area could be estimated and relative abundances in trawlable and untrawlable areas could be estimated). For example, a stratified survey design similar to the NMFS bottom trawl survey design for the Gulf of Alaska may well provide reliable abundance estimates of the portion of the population vulnerable to trawl gear. Given the demersal habits of sablefish, it is not unreasonable to assume that catchability of the trawl is 1 for larger fish residing on the bottom (i.e., all sablefish in the path of the trawl are caught). In the sablefish assessment for the Gulf of Alaska (Hanselman et al. 2006), trawl catches are used as an index of abundance because they only cover the shallow portion of sablefish habitat; however, the index is assumed to reflect absolute biomass within the surveyed area. In the US West Coast sablefish assessment (Schirripa and Colbert 2005) slope survey results are used to estimate absolute biomass of sablefish in different strata, assuming a catchability of 1. Thus, area-swept estimates from a trawl survey in Chatham Strait could be used to estimate the density of sablefish (number per unit area) and to expand the density estimates to an estimate of total sablefish abundance in Chatham Strait.

The major drawback of this approach is that, to my knowledge, a systematic trawl survey of Chatham Strait has not been conducted in the past. Therefore, a new survey would be required and a historical time series to evaluate trends would not be available for some time. Moreover, the costs of an additional survey may be prohibitive, given that it would be highly desirable to maintain several independent estimates of biomass for several years (e.g., tagging study + trawl survey) for calibration. While experience from other areas with trawl surveys suggest that reliable sablefish biomass estimates could be obtained, there would also be some uncertainty as to whether a trawl survey in Chatham Strait would be feasible. Furthermore, the survey would provide a “snapshot” and, like most estimators, would not be able to estimate or account for immigration.

As an alternative to the trawl survey, a relative index of abundance is available from the existing longline survey. The catchability of sablefish in the longline survey could theoretically be estimated to obtain an estimate of absolute abundance. This has in fact been attempted in the past using trawl-longline comparisons to convert a longline index to biomass estimates (Rose 1986, cited in Clausen et al. 1997). However, the results displayed highly unrealistic biomass estimates and the effort was abandoned. The current longline survey does not appear to provide a reliable index of abundance.

Depletion Estimator

Depletion estimators simply evaluate the reduction in some index of abundance as a result of removing (harvesting) fish. The abundance at the beginning of the removal experiment or fishery is estimated based on the relative change in the index caused by removing a known amount of fish. Thus the estimation involves a very simple process that accounts for removals between time t-1 and time t (“Process” model):

$$\text{Abundance}_t = \text{Abundance}_{t-1} - \text{Catch}_t$$

and assumes that the observed catch-per-unit-effort (CPUE) is proportional to abundance (“Observation” model):

$$\text{CPUE}_t = q * \text{Abundance}_t$$

All that is required is a time series of catch and effort (or CPUE). As a simple example, if 1000 tons are removed by a fishing event and a 20% reduction in CPUE is observed, then the original abundance must have been five times as large as the catch or 5000 tons.

This simple model provides an estimate of abundance or biomass at the beginning of the depletion experiment but there is no accounting of any biological processes (growth, R, M, etc.). Potential problems that can bias the estimate include:

- Changes in abundance/biomass for other reasons besides harvesting due to growth, recruitment, mortality, or movement. Therefore the estimator only works over periods that are short enough such that these processes can be ignored.
- The CPUE may not be proportional to abundance due to hyperstability or hyperdepletion. In particular, hyperstability is a common phenomenon as fishers are often skilled at maintaining high catch rates in spite of declining abundances. In such cases, CPUE estimates can give a misleading impression of stable abundances when the population may in fact be declining.

A depletion experiment was conducted in Chatham Strait to estimate sablefish abundances (Clausen et al. 1997), but it resulted in unrealistically high estimates. Clausen et al. (1997) attributed the failure to increases in feeding activity over time (i.e., CPUE did not decrease as much as it should, assuming constant catchability). However, the results could also be due to continuous immigration into Chatham Strait over the period of the experiment. Results from the mark-recapture analyses provide some evidence for this scenario. Depletion estimators are used to assess some species with a short life history. For example, Japanese scientists have used a depletion estimator to estimate fishing mortality and initial abundances of squid in the Japan/East Sea at the beginning of the fishing season (Tian, personal communication[2]).

Surplus Production Model

Surplus production models are similar to a depletion estimator in that catch (C) is “scaled up” to estimate the abundance (N) or biomass (B) that must have been present to create an observed catch series. The process model includes a term for production to account for “bulk” changes in abundance consisting of growth, mortality and recruitment, all of which are lumped into a single term and depend on current stock size ($f[N_t]$):

$$N_{t+1} = N_t + f(N_t) - C_t$$

The observation model is the same as before (CPUE = $q * N_t$). A variety of surplus production models have been developed, which differ only in the shape of the function $f(.)$, in the above equation, for example, the Schaefer, Pella-Tomlinson, and Fletcher models. In the simplest model (Schaefer model), production (change in N from time t to t+1 plus the catch that was removed) is a quadratic function of stock size, while the other models allow for asymmetries in the relationship between production and biomass (Figure 2). While a general production term is included, all the details of growth, recruitment (R) and natural mortality (M) are ignored.

The objective of fitting a surplus production model is finding the maximum sustainable yield and the abundance (or biomass) that maximizes production. Outputs include estimates of average

---

[2] Tian, Y. Japan Sea National Fisheries Research Institute, Fisheries Research Agency, Suidou-cho, Chuo-ku, Niigata 951-8121, Japan.

annual abundances/biomass and estimates of MSY and B_{MSY}. Surplus production models are used in many fisheries around the world, particularly if no age or size-structured data are available. Examples include blue marlin and tuna stocks, assessed by International Commission for the Conservation of Atlantic Tuna (ICCAT), and some small pelagic stocks like sardine and anchovy. Drawbacks of the surplus production model include a lack of biological realism (i.e., oversimplification of dynamics) and poor performance in some simulations (Walters and Martell 2004, p.107). A surplus production model—among other models is also used in management strategy evaluations for sablefish management in British Columbia (Rob Kronlund, personal communication[3]).

Delay-Difference Models

Similar to the surplus production model, delay-difference type models assess overall abundance or biomass dynamics without following individual cohorts over time. These models include recruitment in a given year (R_t) explicitly and keep track of numbers at age (N_t) by assuming annual survival rates that do not depend on age (i.e., $S_{a,t} = S_t$) and knife-edge recruitment:

$$N_{t+1} = S_t N_t + R_t$$

Because typically there exists at least some information on average weight in a fish population, even when no or poor age data have been collected, these models were extended to incorporate information on average weight ($\overline{w} = B_t/N_t$) of fish in a given year. Thus changes in biomass result from processes associated with growth, deaths, and new recruits, which can be modeled as a function of biomass in the previous years, survival from year to year, growth parameters, and new recruits:

$$\begin{aligned} B_{t+1} &= B_t + \text{growth} - \text{deaths} + \text{recruits} \\ &= f(B_t, S_t, \text{growth parameters, recruitment}) \end{aligned}$$

Moreover, to project harvest using this model, a stock-recruitment relationship (e.g., Ricker or Beverton-Holt) must be assumed to model average recruitment as a function of parental biomass.

Important assumptions of the standard (Deriso-Schnute) delay-difference model are therefore:

- Annual survival is the same at each age (i.e., size selectivity cannot be modeled)
- Growth follows a known equation (Brody equation in the original formulation: $w_{a+1} = w_a + \rho\, w_a$, but various extensions/modifications have been used).
- “Knife-edge” recruitment
- Natural mortality is known
- Known stock-recruitment relationship

Delay-difference models offer a compromise between simple, but unrealistic, production models and more complicated age-structured models. Recent applications of such models for stock assessment include prawns in New Zealand (Dichmont et al. 2003), tuna stocks in the Atlantic (ICCAT, http://www.iccat.int/), yellowfin tuna (*Thunnus albacares*) in the eastern tropical

---

[3] Kronlund, A.R., Fisheries and Oceans Canada, Science Branch, Pacific Region, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, British Columbia V9T6N7

Pacific (Quinn and Deriso 1999), and ocean quahog (*Arctica islandica*) and yellowtail flounder (*Limanda ferruginea*) on the east coast of the United States (http://www.nefsc.noaa.gov/).

Simulations have shown that these models perform quite well (almost as well as age-structured models) for stock assessment purposes (Walters and Martell 2004, p.108). However, when reliable age data are available, they have typically been replaced by age-structured models.

Age-Structured Assessment (ASA)

Age-structured assessments incorporate the full age composition, growth, recruitment, maturity, and potentially other processes. Selectivity in the fishery and in surveys is modeled explicitly, therefore survival rates (as determined by natural mortality, fishing mortality, and selectivity-at-age) can vary by age (survivorship: $S_{a,t}$). In modern, statistical age-structured analyses, abundance (and/or biomass) trends may be fit to multiple indices of abundance from independent surveys or from the fishery.

Age-structured models track the number of fish for each cohort over time:

Numbers at age:

$$N_{t+1,a+1} = N_{t,a} \cdot S_{t,a}$$

and calculate the total biomass as a function of numbers and weight-at-age:

Biomass in year *t*:

$$\hat{B}_t = \sum_{a=r}^{A} N_{t,a} \cdot w_{t,a}$$

where weight-at-age may or may not vary over time ($w_{t,a}$ or w_a).

Detailed equations for catch-at-age, age composition ($C_{t,a}/C_t$), size composition, etc., and other model details are described in the section on Age-structured assessment.

Without independent estimates of biomass, this is still basically a depletion estimator, i.e., the relative change in some index of abundance resulting from removing fish is used to “scale up” catches in year *t* (C_t) and reconstruct abundances and biomass (N_t, B_t).

The advantages of age-structured models include increased biological realism that can account for variability in key parameters with age and over time. ASA models have become the model of choice for most demersal fish stocks in the U.S., particularly in the Northeast Pacific, and increasingly elsewhere.

Outputs from an age-structured assessment include:

- Abundance/biomass estimates by year and age
- Fishing mortality estimates by year
- Gear selectivity at age
- Predicted age and size compositions
- Catchability q by gear type (CPUE = q * B)

HISTORICAL AND CURRENT ASSESSMENT APPROACHES FOR CHATHAM STRAIT SABLEFISH

A variety of stock assessment approaches have been tried for the Chatham Strait sablefish population and methods have evolved over time as the quality and quantity of available data has increased (Carlile et al. 2002). Prior to 1988, a relative index of abundance from the fishery was used to set harvest rates. Since 1988, a longline survey has provided a fishery-independent index of abundance. However, no absolute abundance estimate or estimate of exploitation rate to which a harvest control rule could be applied was available until after the tagging program was established and refined (tagging began in 1997). First attempts to develop an absolute measure of abundance included the use of a depletion estimator (Clausen et al. 1997). Failure of the depletion estimator and large uncertainties in the survey-based index of abundance (including an interruption in the time series due to gear changes in 1997), prompted ADF&G to conduct tagging experiments in 1997 and to estimate absolute biomass and/or exploitation rates using capture-recapture methods based on recoveries of fish tagged during the longline survey and recovered in the fishery. Strong evidence that sablefish tagged in the longline survey exhibited gear avoidance (gear shyness), led to a change in the tagging protocol in 2000. Since that time, fish have been tagged and/or marked with fin clips, and released during a pot survey conducted prior to the longline survey and fishery. Estimates of abundance or exploitation rate from these tagging experiments have been used since that time to set quotas for sablefish in Chatham Strait.

Because the longline survey has collected weight and age information in addition to CPUE since 1988, an age-structured model was first fit to the available data in 1995 (based on 1989–1995 data), using a spreadsheet model that was adapted from a model developed for Gulf of Alaska sablefish by Mike Sigler (Auke Bay Lab, AFSC, NOAA) as described in Sigler (1999). The model performed poorly using Chatham Strait data, presumably because of the short time series of available age data, and was felt to result in unrealistically high biomass estimates. An updated ASA in 1998 produced better fits to the data and more plausible results, but was ultimately not used for stock assessment purposes because resulting biomass estimates were still considered to be unrealistically high. The ASA was last updated with 1988–2001 data. Although results seemed reasonable, the use of mark-recapture estimates of abundance was continued to set quotas and the age-structured assessment has not been updated since that time.

OTHER SABLEFISH STOCKS AND OTHER NORTHEAST PACIFIC GROUNDFISH STOCKS

To provide some context I briefly summarize stock assessment approaches currently used for other sablefish stocks and for other groundfish stocks in the Northeast Pacific.

Sablefish Stocks

Sablefish populations along the west coast of North America are currently assessed as three separate stocks (in addition to state-managed populations in Southeast Alaska). Sablefish in federal waters of the Gulf of Alaska, Aleutian Islands, and Eastern Bering Sea are managed as a single stock using a statistical age-structured model. The model uses longline and trawl survey indices, size and age compositions, and independent estimates of natural mortality, weight-at-age, and female maturity-at-age. Since 2007, males and females have been modeled separately using sex-specific weight-at-age relationships. The model is implemented in ADMB and is described in Hanselman et al. (2007).

Sablefish along the US West Coast are assessed using a statistical, age-structured model that is fit to pot and trawl survey indices and to age composition data from surveys and from the fishery. This model is quite unique in that it incorporates two sources of environmental data, sea surface temperature (SST) and sea level height. Sea surface temperature is used to estimate release mortality of discarded sablefish, and sea level height is used to model recruitment deviations from the stock-recruitment function (Schirripa and Colbert 2005). The model is implemented in Stock Synthesis Model 2 (SS2, which uses ADMB for model fitting).

The assessment of sablefish off British Columbia has employed various approaches over time and currently uses an “integrated” tagging model (Haist et al. 2005). Tags are released during pot surveys and the model uses monthly recovery data from the fishery. In addition, various abundance indices are incorporated into the model, which is implemented in ADMB. Management strategy evaluations are conducted using a surplus production model with mark-recapture estimates of abundance from the integrated tagging model.

Other Groundfish Stocks, Northeast Pacific

The vast majority of groundfish stocks in the Northeast Pacific are assessed using statistical age-structured models, including the following stocks grouped by region:

Bering Sea: Walleye pollock (*Theragra chalcogramma*), Pacific cod (*Gadus macrocephalus*), yellowfin sole (*Pleuronectes asper*), Greenland turbot (*Reinhardtius hippoglossoides*), arrowtooth flounder (*Atheresthes stomias*), northern rock sole (*Lepidopsetta polyxystra*), flathead sole (*Hippoglossoides elassodon*), Alaska plaice (*Pleuronectes quadrituberculatus*), Pacific ocean perch (*Sebastes alutus*), northern rockfish (*Sebastes polyspinus*), Atka mackerel (*Pleurogrammus monopterygius*)

Gulf of Alaska: Walleye pollock, Pacific cod, arrowtooth flounder, northern rock sole, flathead sole, Dover sole (*Microstomus pacificus*), rex sole (*Errex zachirus*), Pacific ocean perch, Northern rockfish, dusky rockfish (*Sebastes ciliatus*)

British Columbia: Pacific ocean perch, sablefish (assessed with age-structured model that integrated tag information in 2000)

U.S. West Coast: All major stocks (20+ stocks in 2006)

Several crab stock assessments use statistical size-structured models because of the difficulty of aging individuals. These include assessments of the major crab stocks in the Bering Sea and Gulf of Alaska.

To my knowledge, only the assessment of Pacific cod off British Columbia uses a delay-difference model (with an abundance index) and a surplus production model (with an abundance index) is used for the sablefish stock off British Columbia in a management strategy evaluation context, but not for the assessment.

Survey-based estimates of absolute biomass are used to set catch levels for a number of stocks and stock complexes with relatively poor data:

Bering Sea: “other” flatfish complex, shortraker/rougheye complex, “other” rockfish complex, sharks, skates, sculpins, octopus

Gulf of Alaska: shallow water flatfish complex, rougheye rockfish (*Sebastes aleutianus*), shortraker rockfish (*Sebastes borealis*), other slope rockfishes, other pelagic shelf rockfishes, demersal shelf rockfishes, thornyheads (*Sebastolobus altivelus* and *Sebastolobus alascanus*), skates (various species), weathervane scallop (*Patinopecten caurinus*)

A relative abundance index only is used for other stocks including arrowtooth flounder, lingcod, and longspine thornyhead (*Sebastes altivelus*) off British Columbia. What are believed to be conservative quotas, based only on historical catch series, may be set for stocks that lack even a relative abundance index, including squid in the Eastern Bering Sea, and the deep water flatfish complex and Atka mackerel in the Gulf of Alaska.

CHOICE OF STOCK ASSESSMENT MODEL

There are many considerations that may affect the choice of a model or models to use in assessing any fish stock and the Chatham Strait sablefish stock in particular. I loosely categorized the relevant issues as follows and discuss their relevance in the context of the Chatham Strait sablefish assessment:

Scientific-Statistical Considerations

- Choose a model with "optimal" complexity (simplest model that captures main features of data without overfitting)
- Choose a model that is biologically "reasonable"
- Examine diagnostics/goodness-of-fit/residual patterns, etc.

Model selection criteria are designed to choose the most parsimonious model that leads to the best prediction. However, most formal selection criteria only compare models that are fit to the same data set, thus models that use different sets of data cannot be compared easily. Most commonly, stock assessment models are compared directly in terms of the likelihood for those data components that are used by all of the models being compared. More formally, a retrospective analysis (a form of cross-validation) is typically used to compare the performance of different models over time (i.e., models are fit to progressively more years of data and are examined for biases and variability with respect to the model that includes all years of data).

The simplest models often do not use all of the available data and may lack biological realism. However, any available data may be used in other ways (outside the model) to inform the assessment or the choice of a model. While all models lack complete biological realism, the structure of any model should always be based on reasonable biological assumptions.

Ultimately, the choice of a model is often subjective and based on pragmatic considerations, but any model that is chosen should provide a reasonable fit to any datasets that are deemed reliable and are included in the analysis. Therefore, examining model diagnostics is a critical step in evaluating alternative models.

In evaluating different mark-recapture models and the ASA model, I evaluated model fit based on standard diagnostics such as residual trends and compared models based on the likelihood or on the Akaike's Information Criteria (AIC) model selection criterion, where appropriate. Because the mark-recapture estimates and the ASA model use very different data sets, the model fits cannot be compared directly but the models can only be assessed on their respective merits.

Policy Considerations

- Choose a model that is likely to provide the "best" policy advice and/or simulate performance to search for best policy options.
- Choose a model that allows exploration of different policy options through simulations/management strategy evaluation.

While it would be ideal to conduct management strategy evaluations for each assessment, this is often time consuming and may not be feasible with available resources. Therefore, the performance of a given model has to be evaluated based on generic simulations and accumulated experience. Performance of a variety of models has been evaluated through simulations with "known" scenarios. These simulations have shown that, in the hands of experienced stock assessment scientists, more complex models are likely to perform better (National Research Council 1998, Walters and Martell 2004).

Data Availability

- Choose a model that is supported by and makes best use of the available data

While the availability of data should not necessarily drive the choice of model, a model that is supported by and makes best use of the available data is generally preferable. This would argue for the most "integrated" model that incorporates as much of the available information as possible into a single modeling framework. Therefore, when reliable age data are available, an age-structured model that also incorporates auxiliary data such as abundance indices, is typically the preferred model. The complexity of stock assessment models typically increases over time as the quality and quantity of data improves.

History of Fishery and Management

- Choose a model that is transparent and is likely to be accepted by fishermen and managers.
- Choose a model that captures important changes in fishery and management.
- Strive for continuity and gradual changes.

This argues for a careful, "go slow" approach and a gradual transition to whatever model will be adopted, but also for enough complexity to capture changes in the fishery (for example gear changes, i.e., changes in selectivity, should be accounted for if large changes in management occurred). Hence a balance between complexity and simplicity (which enhances transparency) has to be found. I believe that transparency of the process and continuous dialogue with the stakeholders are essential to achieve buy-in and to be successful in the long run (in particular when the stock undergoes a period of decline, as is currently the case for Chatham Strait sablefish).

Biological Realism

- Choose a model that captures important biological processes (growth, recruitment)

The desire for biological realism generally tends to favor more complex models, but requires careful model development. For example, as shown elsewhere, male and female Chatham Strait sablefish grow at very different rates and the proportion of females in the catch decreases with age. The former suggests that biomass estimates should be obtained by modeling male and female biomass-at-age separately, using different growth curves (weight-at-age), while the latter suggests that female mortality may be higher, that older females emigrate from Chatham Strait, or that selection of bigger fish in the fishery tends to alter the sex ratio at older ages because females tend to be larger at age. None of these features is built into the age-structured assessment presented here. However, female weight-at-age is used to estimate spawning biomass and the use of an average weight-at-age relationship may well be adequate to model the combined biomass dynamics of the population.

The question of how much biological realism is "enough" is difficult to assess and ultimately requires management strategy evaluations to compare the performance of different models. In my view, the age-structured assessment developed here has an appropriate level of biological realism based on the available data, but some of the assumptions with regard to sex ratios, length-weight relationships, or average weight-at-age of males and females should be examined regularly and updated as needed.

Stock Structure

- Choose a model that reflects stock structure

The assessment model should adequately account for the distribution and spatial structure of the population. The current assessment and the age-structured model that I propose assess the population in Chatham Strait as a single unit. Given the spatial extent of Chatham Strait and the degree of movement observed between different statistical areas, it is very unlikely that there is stock structure on a finer spatial scale than the entire Strait. However, there is strong evidence that there is considerable movement of sablefish between Chatham Strait and the outside, suggesting that only a portion of the stock resides inside Chatham Strait at any given time or there is considerable exchange with the Gulf of Alaska stock (presumably a single stock). It is possible that sablefish in Chatham Strait are part of the same stock, which suggests that an assessment model that includes both offshore and inshore areas may best reflect the stock structure of sablefish in the Gulf of Alaska and Bering Sea. Therefore, in the long run, a joint effort with National Oceanic and Atmospheric Administration Fisheries (NOAA) to model sablefish across the entire Gulf of Alaska and Bering Sea, including inshore areas, as a single stock with movement between different areas should be considered. The development of an age-structured model that is very similar to the model currently used by NOAA for sablefish in offshore waters could provide an important step in that direction. Ultimately, this may also include sablefish off northern British Columbia, which are likely to be part of the same stock. However, the Canadian Department of Fisheries and Oceans (DFO) currently uses very different stock assessment approaches that cannot easily be combined with the age-structured analysis proposed here for Chatham Strait and used by NOAA for offshore waters.

A combined stock assessment is unlikely to occur in the near term and the proposed age-structured assessment model for Chatham Strait should ideally account for the movement of sablefish between Chatham Strait and outside waters. Unfortunately, relatively little is known about the extent and nature of the exchanges. An examination of tag releases and recoveries by NOAA, ADF&G, and DFO could shed some light on this and would be very useful in evaluating the validity of the proposed model.

Even with movements into and out of Chatham Strait occurring, results from an age-structured assessment may well provide a reasonable estimate of the total population size, if the Chatham population is distinct from the Gulf of Alaska population. Depending on what form the movements between Chatham Strait and outside waters take, the ASA could still provide an estimate of the total population size (unless forced to fit the mark-recapture abundances). Two possible scenarios are:

- Young sablefish hatched inside Chatham Strait leave the Strait, spend some time outside Chatham Strait and gradually return at older ages (e.g., as they approach maturity). Depending on the pattern of returns with age, this could be reflected in selectivity patterns, i.e., younger fish are simply not available to the gear – the model would not be able to distinguish between size selectivity and the absence of smaller fish from the Strait and would provide an estimate of total abundance. However, the pattern of returning fish by age would have to be fairly regular because the assumed logistic selectivity curve is not very flexible and could not account for a pulse of fish returning to the Strait at a specific age.

- The Chatham Strait population is a distinct population, but individual fish may leave and re-enter the Strait periodically or randomly, such that at any given time only a certain proportion of the population resides inside Chatham Strait. Again, depending on the patterns of emigration and immigration by age (e.g., do fish of all ages have the same migration rates?), the ASA could provide an estimate of total population abundance, including the fish outside Chatham Strait that may not be available to the fishery (but would be available to other fisheries in outside waters). In this case, the missing fish (i.e., fish outside Chatham Strait) would be reflected in the model fit in a lower estimate of the catchability coefficient relating total abundance to the mark-recapture estimates (or to other abundance indices). However, it is unlikely that the model would be able to estimate this parameter reliably and the fraction of fish inside Chatham Strait would probably have to be estimated independently (as, for example, in the case of Greenland Turbot in the eastern Bering Sea, where the survey area is assumed to encompass 75% of the stock). Moreover, if there is substantial fishing mortality on the portion of the stock outside Chatham Strait, these removals should be accounted for, again suggesting a need for a combined model of sablefish in the Gulf and in Chatham Strait with migration between the two or more areas.

Of course, there are many other possible patterns of migration that could lead to biased estimates of abundance.

OVERALL RECOMMENDATIONS FOR GENERAL ASSESSMENT APPROACH

1) With regards to the overall assessment approach I recommend continuing the development of the age-structured assessment model to fully integrate the available data into a single model that can serve to reconstruct historical abundances, assess the current status of the stock, and estimate management parameters, including fishing mortality rate parameters and allowable catches under various fishing mortalities. Most importantly, the model provides an estimate of current stock size relative to the unfished stock size (= depletion). While there are still unresolved issues about the immigration of sablefish into Chatham Strait, the age-structured model may well provide a reasonable estimate of stock size inside Chatham Strait at the time of the longline survey/fishery, if tuned to the mark-recapture estimates of abundance under certain assumptions. The impact of potential immigration during the fishing season (which affects mark-recapture estimates) and throughout the year on model estimates should be examined in the future and could be incorporated into the model as necessary. However, understanding their impact on model estimates will require a better understanding of migration patterns. The model, for the time being, should be fit to a fishery-independent

estimate of abundance (i.e., the mark-recapture estimate), but would replace the current yield-per-recruit analysis for determining $F_{40\%}$ and catch levels.

2) Instead of the yield-per-recruit analysis, $F_{40\%}$ (the fishing mortality that would reduce spawning stock biomass to 40% of the "unfished" level, which is currently used as a reasonably conservative fishing mortality rate) should be estimated within the age-structured model to determine the allowable catch corresponding to the harvest control rule.

3) At least for the first several years, until more experience can be gained with the age-structured model and some of the migration issues can be resolved, I recommend continuing the tagging program to obtain an independent estimate of absolute sablefish abundance in Chatham Strait using the Petersen estimator or one of the suggested variants. This estimate of abundance will serve to constrain ("anchor") the biomass estimate in the age-structured analysis, which may be necessary as long as the absolute level of biomass estimated by the model is highly uncertain and sensitive to small changes in the model. As the ASA is improved to reflect the dynamics of the Chatham Strait population more accurately, independent ASA estimates (i.e., estimates not tuned to mark-recapture data) and either mark-recapture estimates will converge or reasons for any discrepancies will be better understood.

4) If the age-structured model is adopted and provides reasonable estimates of past abundance trajectories and management parameters, the importance of the tagging data, longline CPUE data, and fishery data to the model should be re-evaluated periodically. Sablefish are long-lived with moderate recruitment variability and their biomass is not expected to change dramatically over time. Therefore, it may be sufficient to conduct the pot survey, the longline survey, or both on a biennial basis to monitor the Chatham Strait stock. Simulations with the age-structured assessment could be conducted to evaluate the associated risk, which is likely to be relatively low. For example, stock assessments of a number of important groundfish stocks in the Gulf of Alaska (including flathead sole, rex sole, other flatfish, Pacific Ocean perch, northern rockfish, rougheye rockfish, and thornyheads) are conducted on a biennial basis to coincide with new trawl surveys.

5) In the long term, continued coordination with the NOAA and DFO sablefish assessments would be desirable because there is evidence that sablefish in Chatham Strait are part of a single sablefish population ranging from at least British Columbia into the Bering Sea. It is conceivable that over time a joint assessment could be developed, at least between NOAA and ADF&G, and that area apportionments could be used to determine catch quotas for Chatham Strait and offshore areas. Coordination with and integration with the NOAA assessment would be facilitated if an age-structured model is developed for Chatham Strait that is similar to the one used by NOAA. The model developed here is a step in that direction but would require a split by sex in the future.

6) Gradual improvements to the age-structured assessment that could be explored as time permits include:

a) Refinement of CPUE indices, for example using a General Linear Model to account for the influence of different covariates (e.g., vessel, sampling date, area, etc.) on CPUE (Hilborn and Walters 1992).

b) Use of a split-sex model to improve model fits due to large differences in growth between males and females. The possibility that natural mortality differs between males and females should also be explored.

c) Inclusion of a stock-recruitment estimator into the age-structured model to provide direct estimates of F_{MSY}, which could be used instead of a relatively arbitrary proxy such as $F_{40\%}$.

d) Verification and modification of the age-transition matrix as necessary based on data from the ADF&G aging lab because the current matrix is based on data from the NMFS aging lab.

e) Exploration of alternative selectivity curves with additional flexibility.

f) Periodic re-examination of growth and maturity schedules.

g) In the long-term, development (possibly in conjunction with NMFS) of a spatially explicit model that includes migration.

7) The National Research Council has provided a very useful checklist for stock assessment scientists:
http://www.nap.edu/openbook.php?record_id=5951&page=137

MARK-RECAPTURE ESTIMATES OF SABLEFISH ABUNDANCE IN CHATHAM STRAIT

ASSUMPTIONS AND THINGS THAT MAY GO WRONG

The underlying assumptions of a mark-recapture estimate can be violated in various ways, but the consequences of such violations are not always obvious. Here I examine violations in terms of each of the quantities in the fundamental equation (M, N, m, n):

Changes in *M*

The number of marked fish (M) in Chatham Strait that are available to the fishery is assumed to remain constant but may change between the tagging event and re-capture event due to violations of the closure assumptions or tag loss. Marked fish may die (natural mortality), leave the area (emigration), or lose their external tags. The latter is less of a concern because all marked fish are also clipped, which allows an estimation of the tag loss rate, and there is no loss of marks when only clips are considered.

Uncertainty/Bias in *n*

The sample collected for the “recapture event” (n) may not be a random subsample from the fishery, thus it may differ in important aspects from the overall population. This could arise from gear differences between the fishery (longline gear) and the tagging event (pot gear) or from sampling issues associated with the fishery samples. The assumption of equal capture probability implies, among other things, that the re-sampled population (i.e., the population exploited by the fishery) is the same as the tagged population and that the samples are random samples from the population. Clearly there are differences between the sampled and re-sampled populations, particularly in terms of their size composition, that may affect estimates.

Uncertainty/Bias in *m*

The number or proportion of recovered marks in the fishery sample may be biased for a number of reasons.

- Fish are subject to fishery selectivity and tagged fish may be under-reported or discarded.
- Tagged fish may be retained disproportionately. For example, while many small fish are discarded, small fish with tags appear to be retained in proportion to their abundance.
- Tags or clips may be overlooked during sampling
- Tagged fish may have a different probability of recapture (e.g., gear shyness, lack of mixing).

Uncertainty/Bias in *N*

The population of interest (*N*) may be defined in different ways and different estimators estimate different "populations". In particular, there is pronounced size structure in the population from small fish that are not caught by conventional gear at all, or have very low gear selectivity, to large, fully selected individuals. Therefore, it is not always clear what the "vulnerable" population is that is being estimated. This is particularly important here because a different gear with a different selectivity is used during the capture event (pot gear) than during the re-capture events (longline: either survey or fishery). Because of differences in gear selectivity between pot and longline gear and between the longline survey and the fishery, the "vulnerable" or "exploited" population can be defined in different ways and different estimators estimate different vulnerable populations. We explored this and a number of other issues through simulations described below.

Furthermore, the population of interest (*N*) may also change between tagging and re-capture through immigration, emigration, mortality, and recruitment (violation of closure assumption).

CHATHAM STRAIT TAGGING PROGRAM

The Chatham Strait tagging program currently uses a pot survey to capture and release fish that are marked with tags and fin clips or fin clips only. Marked fish are subsequently recovered in the longline fishery as well as in the longline survey (taking place prior to the fishery). A considerable amount of effort has gone into improving and refining the tagging program over time. Several years of relatively standardized marking and sampling are now available to estimate the abundance of sablefish in Chatham Strait. The current approach and several alternatives for estimating abundances from the tagging program are described in detail below.

Possibly the largest advantage of the mark-recapture approach is the extensive experience with tagging sablefish in Chatham Strait that has been accumulated over the years. The experience gained has resulted in reasonable and quite consistent estimates of abundance since at least 2003. While there may be a number of potential problems that remain to be resolved, modest adjustments to the sampling program and some refinement of the modeling approach should be able to provide reliable estimates of absolute abundance. One of the reasons for re-evaluating the mark-recapture approach is the unrealistically tight confidence interval resulting from the modified Petersen estimator. One of the goals of my analyses was therefore to develop an approach that results in more realistic confidence intervals.

ISSUES EXAMINED

- Natural mortality
- Emigration
- Immigration
- Difference in mortality between marked & unmarked fish
- Difference in probability of capture
    - Sorting/size-selectivity issues
    - Spatial patterns

I evaluated the potential importance of each of these issues and, to the extent possible, incorporated presumed mechanisms in models. I fit a number of models incorporating some of the mechanisms that are potentially important.

Natural Mortality

Substantial natural mortality may occur between tagging and recapture. In recent years, a substantial amount of time passed between the marking phase (pot survey) and the recovery phase (fishery). For example, the average marking date (mean date of pot survey) in 2006 was June 8, which was 70 days prior to the start of the fishery. The annual instantaneous mortality rate of sablefish is assumed to be approximately 0.1 in the stock assessment. Assuming that mortality occurs throughout the year (the usual assumption in stock assessments), this equates to a daily instantaneous mortality rate of $m = 0.1/365 = 0.000274$, and the number of "surviving" marks after t days is: $M_t = M_1 \exp(-m*t)$. To account for natural mortality, I assumed a daily instantaneous mortality rate of 0.000274 in the model for both tagged and untagged fish. If handling mortality is significant and can be estimated it could be included by assuming different instantaneous mortality rates for marked and for unmarked fish. Under the above assumption, the number of tags available in Chatham Strait at the end of the fishery (November 15) was 5,331 instead of 5,582 (the latter number accounts for all known removals but ignores natural mortality (m) (Figure 3).

The assumed value for natural mortality itself is uncertain, but small variability in m had little effect on the results. Because it is straightforward to account for natural mortality in the estimation it seems prudent to do so.

Emigration

Emigration out of Chatham Strait is known to occur and reduces the number of marked fish (M) that are available to the fishery. Total emigration from the time of marking to the end of the fishery can be estimated from the number of tagged fish that are recovered outside Chatham Strait. The current approach accounts for known emigrants only by subtracting tags returned from outside Chatham Strait from the number of tagged fish. This underestimates the number of fish emigrating because only a fraction of the emigrants are caught and some tags may be caught but not reported. The number of reported tags recovered outside Chatham Strait depends on the number of tags released, the probability of capture (exploitation rate) and the tag return rate.

Depending on the outside exploitation rate (as well as the tag reporting rate), for each tag returned from outside Chatham Strait 9 or 10 marked fish may have emigrated but were not caught (assuming an "outside" exploitation rate of about 0.1, as estimated by Hanselman et al. 2006 for the Gulf of Alaska sablefish stock). Emigration can be estimated based on reasonable estimates of reporting rates and exploitation rates. If both the reporting rate and exploitation rate are the same inside and outside Chatham Strait, the emigration rate is simply the number of tags returned from outside divided by the total number of returned tags. Otherwise, the reported number of returns can be adjusted for differences in reporting rates and differences in exploitation rates between Chatham Strait and outside waters (Alaska and B.C.). The adjusted number of tag returns can be computed as described in Haist et al. (2002):

$$R_{outside,adjusted} = R_{outside} \cdot \frac{\rho_{\text{Chatham}}}{\rho_{\text{outside}}} \cdot \frac{u_{Chatham}}{u_{outside}}$$

where: $R_{outside,adjusted}$ = adjusted outside tag returns
$R_{outside}$ = reported outside tag returns
$\rho_{Chatham}$ = Chatham reporting rate
$\rho_{outside}$ = outside reporting rate
$u_{Chatham}$ = Chatham exploitation rate
$u_{outside}$ = outside exploitation rate

The reporting rate in Chatham Strait can be estimated from the number of clipped fish observed in the fishery (only those landings that were completely observed) and the number of tags returned from those same landings (the 2006 Chatham Strait reporting rate was = 356/425 = 0.838). For lack of better information, I assumed that the reporting rate in outside water is the same. Haist et al. (2002) provide some estimates of Alaska and B.C. tag reporting rates that may or may not be applicable in 2006.

The exploitation rate in outside waters (East Yakutat and Southeast Alaska, Tables 3.1a and 3.11 in Hanselman et al. 2006) was 0.0989, which is almost identical to the estimated exploitation rate in Chatham Strait (0.0985 in 2006). I crudely estimated the latter from the number of clipped fish caught in the 2006 fishery (598, adjusted by tag reporting rate) divided by the total number of clips released.

Because the ratios of the Chatham and outside tag return rates ($\rho_{Chatham}/\rho_{outside}$) and of the corresponding exploitation rates ($u_{Chatham}/u_{outside}$) are 1 or close to 1, the adjusted tag return (4.975) is almost identical to the reported return. Assuming that the "outside" exploitation rate ($u_{outside}$) applies to the tagged fish that emigrated, the total estimated number of emigrants is $R_{outside,adjusted}/u_{outside}$ = 4.975/0.0989 or approximately 50 fish, which implies an emigration rate of 0.0094 (50/5325, the ratio of the number of emigrated fish to the total number of tags released), or approximately 1% between the time of the pot survey and the time when the tag returns are reported. Most likely, the outside exploitation rate during the limited time period between the survey and the tag reporting date (u') is considerably smaller than the annual exploitation rate ($u_{outside}$). Therefore the total number of tagged fish that emigrated is $R_{outside,adjusted}/u'$, which is much larger than 50, hence the actual emigration rate is likely to be much larger than 1%.

The estimated emigration rate should be added to the natural mortality estimate that is applied to marked fish to adjust the number of marks in the population in the mark-recapture models. For example, if the emigration rate is assumed to apply to a 100-day period between the pot survey and the approximate mid-point of the fishery and is added to natural mortality over this time period, the estimated rate of mortality + emigration on an annual basis is 0.136.

The effect of emigration is identical to that of natural mortality. Viewed as an instantaneous rate of emigration, the estimated emigration over the course of the season (~ 100 days from time of marking to end of fishing season) is equivalent to an increase in natural mortality from 0.1 to 0.136 (on an annual scale). In the mark-recapture models for 2006 I had assumed a much smaller emigration rate (< 0.1%, rather than 1%, due to a mistake in the calculations) and based on simulations over a limited range of natural mortality values I concluded that emigration was negligible compared to natural mortality. I therefore assumed a natural mortality rate of 10% in all models. Given the revised estimates of emigration and the fact that they are likely to underestimate actual emigration rates, emigration may be more important than I had assumed. Therefore I recommend that the rate of emigration between the pot survey and the fishery be re-examined as part of a study of sablefish movements between Chatham Strait and outside waters.

Immigration

There is some evidence that substantial immigration into Chatham Strait occurs based on tagging studies (Maloney and Heifetz 1997) and based on the 2006 fishery data, which show a significant decrease in the ratio of marked to unmarked fish over time (Figure 4).The decrease could result from immigration of unmarked fish into Chatham Strait (thereby “diluting” the number of marks in the population), higher mortality rates for marked fish, or a combination of both (or effects of limited mixing, or other factors). The effect on estimates of abundance is a steady increase in the running Petersen estimator (which incorporates cumulatively more data into the estimate by adding daily marked and unmarked catches), as well as in a time-stratified Petersen estimator, which suggests a steady increase in estimated biomass within Chatham Strait (Figure 5). A model that includes and estimates a daily constant immigration rate suggests a substantial increase in the number of sablefish over the season, adding greater than 10,000 sablefish per day (Figure 6). However, such a large increase in abundance is not evident in an increase in catch per unit effort (CPUE) in the fishery over time. In fact, average CPUE decreases over time for both conventional gear and snap-on gear (Figure 7) and all statistical areas (Figure 8), which was the basis for depletion experiments in the past. A generalized modeling approach (GAM and GLMs with delivery day, stat area, vessel, and gear type as explanatory variables and CPUE or fourth-root transformed CPUE ($CPUE^{0.25}$) as response variable) shows a significant non-linear decrease in CPUE over time with no significant differences among statistical areas, vessels, or gear types. The “best” model was a regression of CPUE on delivery day (Figure 9). Conclusions were the same regardless of whether raw CPUEs or fourth-root transformed CPUEs were used (to reduce influence of some large outliers). Therefore the fishery CPUE data does not provide evidence for substantial immigration of sablefish into the Chatham Strait area over the course of the fishing season. CPUE also appeared to be affected by the total number of hooks fished and tended to decrease as the number of hooks increased (Figure 10). Thus there appears to be some evidence of saturation, i.e., the number of sablefish caught does not increase linearly with the number of hooks (Figure 11).

Difference in Mortality Between Marked and Unmarked Fish

One obvious alternative that could explain a decrease in the ratio of marked to unmarked fish is that marked sablefish experience higher mortalities than unmarked fish. To examine this possibility, I constructed a model that incorporates differential mortality between marked and unmarked fish. The best model fit implies a substantially higher mortality rate of marked vs. unmarked fish (20 times as high for marked fish). This seems highly unrealistic, given that in previous years a large number of tags were recovered in subsequent years. If mortality rates of marked fish were this high throughout the year, only 426 marked fish would be expected to survive until the following year (Aug 4, 2007). For the 2001 to 2005 release years, the number of tagged fish that were recovered in the year following tagging averaged 80% (ranging from ~70 to ~90%) of the number of tagged fish recovered during the year of tagging. Therefore the number of tagged fish recovered in 2007 is expected to be 429 fish, which is similar to the total number of marked fish that is expected to be present in the population under the above model. Therefore the model is clearly unrealistic.

Differences in Probability of Capture

Sorting/Size Selectivity

Another alternative that could explain the decreasing ratio of marked to unmarked fish is a time trend in the relative discarding rates of marked and unmarked fish over the course of the season. Fishermen are likely to discard smaller fish and may or may not discard marked fish at a higher rate. Regardless of the absolute discarding rates, the observed patterns could imply a trend in the relative, size-selective discarding rates under this model (for example if fishermen are more selective at the beginning of the season and discard small unmarked fish, but become less selective as the season goes on). I therefore examined changes in the size composition over time for possible clues on changes in discarding practices. Average size (assuming that the sablefish selected for length and weight measurements were random subsamples of the catch) does not change significantly over the course of the sampling season (Figure 12, Figure 13). Similarly, the full size distribution does not appear to differ over time (Figure 14) and a Kolmogorov-Smirnov test does not show a significant difference among four consecutive time periods (Figure 15, $p > 0.15$ for all pairwise comparisons).

Because size selectivity has important effects on the probability of capture of fish in the survey and in the fishery, I examined the issue in more detail. There are obvious differences in size selectivity between pot and longline gear, and probably between the longline survey and the fishery, resulting in different length-frequency distributions of sablefish sampled during the surveys and the fishery (Figure 16). Some of the differences may be accounted for by seasonal growth, but seasonal growth alone is very unlikely to account for the large effects. The cumulative size distributions showed no significant differences among statistical areas (Figure 17, Figure 18, K-S test: $p > 0.05$), except among areas 3345701 and 345731 (K-S test: $p = 0.015$). When pooling all length measurements across statistical areas and without any weighting, there was a highly significant difference in the size distribution between the two surveys and between each of the surveys and the fishery (Figure 19, $p < 0.0001$ for each pairwise comparison). The size distribution in the pot survey was significantly shifted to smaller sizes relative to the longline surveys and sablefish sizes in the longline survey were significantly smaller than in the fishery.

The size composition in the pot survey differed significantly among statistical areas, both in terms of average size and in terms of the full size distribution (Figure 20), and increased significantly with depth. The differences can largely be captured by a non-linear trend with depth and with latitude (Figure 21). No significant trends in average length with latitude or depth were evident during the longline survey, although there was a marginally significant difference in average length among sampling strata, assuming that each fish sampled for length is an independent sample (Figure 22, ANOVA: $F = 2.655$, $p = 0.047$).

Because sizes differ by depth in the pot survey (Figure 21), it is particularly important that the depth distribution of the pot survey is not too different from that of the fishery, unless these differences are taken into account. I did not have data on depth distribution in the fishery and it is possible that at least part of the reason that the size composition in the pot survey is shifted to smaller sizes is a result of shallower depths being sampled by the pot survey. For example, the pot survey includes 7 stations with a mean depth <210 fathoms (fm), whereas the longline survey did not include any shallow stations with a mean depth less than 210 fm (Figure 23). However, the influence of the shallow stations on the size composition of marked sablefish is reduced by the fact that relatively few sablefish were caught at the shallowest stations and the largest differences in the depth distribution of sablefish between surveys occurs between 250 and 300 fm. The longline survey caught greater than 40% of all sablefish within this depth range, whereas the pot survey caught less than 20% of the total catch in the same depth range (Figure 24).

One important implication of gear selectivity is that the probability of capture differs between small and large fish. This implies that only a proportion of the smaller fish are effectively available to the longline gear, as determined by size selectivity. This becomes an issue if the pot survey marks a proportionally large number of small fish, but because of size selectivity not all of these marked sablefish are effectively available to the longline gear. This in effect reduces the number of marked fish available to the fishery, thereby reducing Petersen estimates of abundance (smaller "effective M" in the numerator). However, the resulting estimate may still be an unbiased estimate of the exploitable population (the portion of the population that is actually susceptible to exploitation, i.e., to being retained in the fishery). At present, problems associated with size selectivity are being alleviated through choosing a cut-off point for marking fish, i.e., only fish 510 mm or larger were clipped and only fish 560 mm and larger were tagged in 2006. Thus clipped and tagged fish can provide estimates of the number of fish above 510 and 560 mm, respectively. I explored the impacts of differences in size selectivity and ways to estimate (and thereby account for) size selectivity of the different gears through simulations (Appendix A.1. Two important observations regarding size selectivity are: (1) the size composition of recovered (tagged) sablefish in the fishery does NOT differ from the size composition of sablefish tagged and released during the pot survey (Figure 25), and (2) the size composition of recovered (tagged) sablefish in the fishery DOES differ substantially from the size composition of all untagged fish caught in the fishery (Figure 26). These patterns could arise if tagged fish have a different selectivity from untagged fish (regardless of differences in gear selectivity), for example if fishermen keep all tagged fish but discard untagged small fish. The latter has been observed in the B.C. fishery and the tagging model used to assess B.C. sablefish makes adjustments for on-board fish sorting, whereby smaller fish are released if they have no tag. The influences of differences in selectivity among gears and between tagged and untagged fish are more fully explored in Appendix A.1

Spatial Differences in Capture Rates

Another obvious mechanism that could lead to differences in capture probability is spatial differences in capture and recapture rates. For example, if a disproportionate number of sablefish are marked in a specific area and there is limited mixing among areas, few of the marked fish may be recaptured if the fishery has a very different spatial distribution. As a result, estimates of abundance will be biased high (i.e., the population will be overestimated) because the number of recovered tags, *m*, is smaller than expected under complete mixing. To guard against unequal capture probabilities resulting from spatial patterns in the fishery and incomplete mixing, the pot survey has been designed to mimic the spatial distribution of the fishery as closely as possible (i.e., sampling effort is distributed among areas in proportion to fishing effort). This provides the best insurance against potential biases associated with limited mixing. As long as there is good mixing over the area represented by sampling stations and the number of marks is approximately proportional to the anticipated fishery harvest in that area, such biases should be minimized.

Previous analyses have examined within-season movements between strata, which suggest reasonably good mixing of the sablefish population in Chatham Strait. If mixing is limited and capture probabilities differ among areas, a spatially stratified estimator may be more appropriate. However, mixing rates in 2006 were fairly high (Table 1). For example, 37% of all recovered sablefish whose recovery location was known were recovered from a different statistical area than the one they were tagged in. Most of these were recovered in adjacent statistical areas (105 out of 164 = 64%). Because of the relatively high proportion of fish migrating to other areas, mixing within a stratum is very likely to be sufficient to ensure that, on average, marked fish have the same probability of encountering a longline as unmarked fish (but may have a different probability of actually being sampled due to other reasons). Locations of release and recovery also show that the average distance that sablefish moved between marking and recovery far exceeds the average distance between sampling stations.

In summary, capture rates of sablefish in Chatham Strait certainly differ by size and may differ between marked and unmarked fish due to sorting (i.e., selective retention of small tagged fish and regular discarding of non-tagged small fish). Issues due to size selectivity and models to address these issues are therefore explored elsewhere in the report. Spatial differences in capture probabilities associated with spatial patterns in the fishery could also result in biases. However, I believe that this is not a concern at present because tagging effort is spread out in space in proportion to fishery harvest and mixing rates are relatively high. Therefore I did not explore models for unequal capture probability associated with spatial patterns in the distribution of the fishery.

EXTENSIONS OF PETERSEN ESTIMATOR

The modified Petersen, or Chapman, estimator that was used in previous assessments (referred to as Model A hereafter) makes some fairly restrictive assumptions, and there was some concern that the estimator may be biased and has an unrealistically small estimated variance. Bias may arise because some or all of the underlying assumptions may be violated as discussed above and the variance is likely too small because the sources of uncertainty in *M*, *n*, and *m* are ignored. I therefore explored several possible extensions to the Petersen estimator to identify and, if possible, incorporate sources of uncertainty and bias into the estimates. The notation used in the development of models is summarized in Table 2.

INCLUDING NATURAL MORTALITY

As a precautionary measure and to remove a possible source of bias, I included natural mortality between the time fish are marked and the time they are recovered in the fishery. The fishery currently extends over approximately 3 months and the time elapsed between marking (pot survey) and recovery in the fishery is at least two months and can be up to five months. This extended duration implies that the assumption of closure will almost certainly be violated. One obvious violation is through natural mortality (e.g., predation) that is likely to occur throughout this period. Natural mortality reduces the number of marks in the population and will therefore tend to reduce the number of marks recovered. Consequently, the Petersen estimator will tend to overestimate abundance if the number of available marks is not discounted for mortality. While it is largely unknown when natural mortality occurs, as a precautionary approach I assume that it is occurring throughout the year. I therefore included natural mortality in all of the models examined to discount both the number of marks in the population and the total number of sablefish. The best available evidence suggests that annual natural mortality of sablefish is around $m = 0.1$ (Hanselman et al. 2006), which translates into a daily instantaneous mortality of $\mu = 0.1/365 = 0.000274$. As discussed previously, emigration has the same effect as natural mortality by removing both marked and unmarked fish from Chatham Strait. The assumed rate of $m = 0.1$ therefore includes both natural mortality and emigration. Because there is uncertainty about this rate I also allowed for moderate variability in m in each of the models and found that it had little effect on the results. I present results that allow for some variability in m (in the form of a relatively narrow prior distribution with mean 0.1 and standard deviation 0.02, see Figure 27). However, because emigration may be more extensive than previously assumed, updated model runs should probably use a higher rate.

TEMPORAL STRATIFICATION

Because of an apparent and significant trend in the ratio of marked to unmarked fish in the fishery catches over time (Figure 4), I stratified the Petersen estimator by time period in each of the following models. To stratify by time, I divided the fishing season into 5-day periods based on date of delivery for a total of 18 periods or strata, denoted by subscript i. The longline survey was treated equivalently as a single stratum. I then computed the number of fish examined for marks (n_i) and the number of marks recovered by period (m_i). The number of marks available at the beginning of each period (M_i) was estimated based on the initial number of marks released (minus known removals from outside Chatham Strait or other fisheries) and the assumed natural mortality rate. Using m_i, n_i, and M_i, the usual Petersen estimator could be computed separately for each time period (Figure 5) and averaged across time periods. As expected from the observed pattern in the ratio of marked to unmarked fish, the stratified Petersen estimates increase over time. A major goal of the models described below is to capture and account for this apparent increasing trend. To be able to compare different models and describe uncertainties in the parameter estimates, I fit the models using both a likelihood-based approach (for model comparisons) and a Bayesian approach that incorporated prior information and additional parameter uncertainty.

The first extension of the Petersen estimator therefore consists of a time-stratified estimator of abundance (abundance at the beginning of the fishery, N_1, or average abundance, $\overline{N}$) that accounts for natural mortality (Model B). Parameters of model B include initial abundance N_1 and natural mortality m (which is estimated independently and highly constrained in the model).

INCLUDING IMMIGRATION

The next model extension (Model C) included immigration into Chatham Strait. Immigration increases the number of unmarked fish in Chatham Strait and could therefore account for the observed decrease in the ratio of marked to unmarked sablefish (Figure 4). There is some evidence that substantial immigration into Chatham Strait may occur based on tagging studies (Maloney and Heifetz 1997). Immigration was included in this model by assuming a daily constant immigration rate β (number of fish entering Chatham Strait per day). This increases the modeled population size at the beginning of each fishery period (N_i) but does not affect the number of marks (M_i). Parameters of model C are average abundance ($\overline{N}$), daily immigration (β), and natural mortality (m). As before, m was narrowly constrained around 0.1, while $\overline{N}$ and β were freely estimated by the model. I estimated $\overline{N}$ instead of N_1 in this model because immigration increases the number of fish available to the fishery. If the fishing quota were based on N_1, fishermen would essentially be unable to exploit the portion of the population that enters Chatham Strait after the beginning of the fishery. Using average abundance (averaged across the 18 fishing periods) provides a more robust measure of the exploitable population because it is less sensitive to variations in the estimated immigration rate than either initial abundance (N_1) or final abundance (N_{18}) and because it is comparable to the Petersen estimator (which averages across all time periods by pooling data). As an alternative, initial or final abundance could be used in all cases.

INCLUDING FISHERY CPUE DATA

As a further extension to the above models and to better approximate the apparent seasonal trend in abundance, I included fishery CPUE as auxiliary data in the model. While models with immigration suggest an increase in abundance over time, the fishery CPUE values suggest a stable or decreasing trend in abundance (density) over time (Figure 28). I computed fishery CPUE (number of sablefish per 1000 standardized hooks) for each trip (or portions of a trip with separate weight measurements) from fish ticket data. For trips that had no sablefish counts, I converted total catch weights to numerical abundance by dividing total catch (in lbs) by the mean weight of a fish across the season. I then computed the average CPUE by stratum (5-day period) across all trips delivered during the 5-day period.

To utilize the information available in the fishery CPUE data I constructed versions of model B and model C that included the fishery CPUE data. I assumed that fishery CPUE was proportional to total sablefish abundance in each time period ($CPUE_i = q * N_i$) and that the difference between the observed and the model-estimated CPUE followed a normal distribution. The estimated time trend in abundance was then tuned to the fishery data by maximizing the normal likelihood (equivalent to minimizing the sum of squared differences between the observed and estimated $CPUE_i$).

The models including fishery CPUE data were fit to both the mark-recapture and fishery CPUE data by maximizing the combined likelihood (consisting of a binomial likelihood component for the mark-recapture data and a normal likelihood for the fishery CPUE data). Both likelihood components received equal weights in the combined likelihood, thus fishery CPUE and mark-recapture data contribute equally to the parameter estimation. The two additional models including fishery CPUE data were equivalent to models B and C, but include the fishery CPUE data which requires the estimation of two additional parameters corresponding to the catchability coefficient q and the standard deviation (σ) of the CPUE data. Model D includes no immigration and estimates a total of 4 parameters (N_1, q, σ, m), where m is highly constrained (Figure 27). Model E extends model D by also including a parameter for daily immigration (β). Essential characteristics of models A–E are compared in Table 3.

INCLUDING EFFECTS OF FISHERY SELECTIVITY

To account for size selectivity of the fishery, which reduces the number of marked (at least clipped) fish that are effectively available to the fishery, I computed a correction factor to adjust for the proportion of marked fish that are available. Simple simulations show that the correction factor depends only on the selectivity of the fishery (if all marked fish in a size class have the same selectivity) and on the number of marks in the population by size class Appendix A.1. For each size class, the number of fish that are vulnerable to fishing can be computed as the product of selectivity times the number of marks in that size class. Summing the number of vulnerable marks across size classes and dividing by the total number of marks released into the population gives the desired correction factor (proportion of marked fish available to fishery). Because selectivity by size has not been estimated for Chatham Strait, I computed fishery selectivity at size based on size-at-age and selectivity-at-age information from the federally managed IFQ fishery as summarized in Hanselman et al (2006). The fishery uses the same longline gear and has been assumed to have a similar selectivity to the Chatham Strait fishery. Average size-at-age data were converted to average age at size, which was then used to compute selectivity by size class (center of size class) based on the average age in this size class. I did these simulations prior to completing the ASA model, which may provide better estimates of selectivity at age for Chatham Strait that could be used instead of NOAA selectivities.

The number of clipped fish in each size class and the estimated selectivity at size are summarized in Table 5. Summing the proportion of clipped fish available to the fishery across size classes and dividing by the total number of clipped fish resulted in a correction factor of 0.748. Therefore, approximately 75% of clipped fish are available to the fishery if both marked and unmarked fish have the same selectivity pattern. However, size composition data suggest that clipped fish have a higher selectivity than unclipped fish in a given size, particularly at smaller sizes, suggesting that clipped fish are retained while other small sablefish tend to be discarded. These and other selectivity patterns are further explored in the Appendix A.1.

DESCRIPTION OF ALTERNATIVE MODELS

I fit a total of five alternative models that extended the simple Petersen model by including time strata, natural mortality, immigration, and auxiliary CPUE data from the fishery. Model A of the assessment corresponds to the Petersen estimate used in previous years. Models B–D are extensions of the Petersen estimate based on the same underlying binomial model and are all stratified by time. Strata consisted of the longline survey and 18 consecutive 5-day fishery periods (based on delivery date). For each time period *i,* the number of recovered (clipped) fish was assumed to follow a binomial distribution:

$$m_i \sim Binom(n_i, p_i)$$

where n_i is the number of fish examined for marks in period i and p_i is the probability of a fish being clipped. The probability p_i is equal to the proportion of the number of marks M_i in the population during a given time period: $p_i = M_i/N_i$ and both M_i and N_i are modeled as functions of initial numbers (N_I, M_I), natural mortality, and immigration. Parameter estimates were obtained in two ways. First, by maximizing the binomial likelihood over all periods and second, the full posterior distributions of each parameter was estimated using a Bayesian model formulation implemented in WinBUGS.

The first time period for the stratified estimator was the longline survey, which was considered to have occurred on a single day (August 4, the average survey date for 2006). The fishery was divided into 18 consecutive 5-day periods for this analysis, based on delivery dates, to ensure that a minimum of 5 marked sablefish were recovered in each period. The date associated with each fishing period was the first day of the 5-day period, which roughly corresponds to the average day of capture (considering that fish are delivered ~1–4 days after being caught).

Based on ADF&G catch data the following quantities were computed for each time period:

- total catch (C_i)
- number of fish examined for clips (n_i)
- number of clips that were found (m_i) in each sample of size n_i, and
- number of days elapsed since the previous time period (t_i).

Based on these data and the model parameters, other quantities were computed as follows:

1. To account for natural mortality of marked fish between the time of the pot survey (marking) and the first period (longline survey):

$$M_1 = M_0 \exp(-\mu d) - D \tag{1}$$

where M_0 is the initial number of marks released, d is the number of days between the middle of the pot survey and the longline survey, D is the number of known removals prior to the longline survey or from outside Chatham Strait, and μ is the daily instantaneous mortality rate, which is 1/365 of the annual instantaneous mortality rate m.

2. To estimate the total number of sablefish and the number of marked sablefish in Chatham Strait at the beginning of each time period:

$$\begin{aligned} N_i &= \left(N_{i-1} - C_{i-1}\right)\exp(-\mu t_i) + \beta t_i \\ M_i &= \left(M_{i-1} - m_{i-1}\right)\exp(-\mu t_i) \end{aligned} \tag{2}$$

where M_1 is from Eq. 1, N_1 is a parameter of the model to be estimated, C_i is the total number of sablefish removed during period i, and m_i is the total number of marked sablefish removed during period i, and β is the number of sablefish immigrating into Chatham Strait per day.

3. Likelihood function (Mark-recapture component of likelihood): The probability of finding m_i marked fish during period i, given the proportion of marked fish in the population (p_i) and the number of fish examined for marks (n_i), follows a binomial distribution:

$$L_i(m_i \mid p_i, n_i) = \binom{n_i}{m_i} p_i^{n_i} (1 - p_i)^{n_i - m_i} \tag{3}$$

The proportion $p_i = M_i/N_i$ depends on parameters N_1, μ, and β (through N_i , Eq. 2) and on the number of marks M_i, which is a function of the initial number of marks and natural mortality (and is not affected by immigration). Assuming independence among successive "sampling" periods (5-day intervals), the total likelihood is a product of the individual likelihoods for each of the 19 time periods:

$$L = \prod_{i=1}^{19} L_i$$

In addition to obtaining maximum likelihood estimates and likelihood-based statistics for model comparisons, I also fit each model in a Bayesian context using WinBUGS. The Bayesian posterior distribution is simply the product of the likelihoods and the prior distributions. In most cases, vague priors were used, except for the natural mortality rate m.

Essential aspects of models B-D are summarized below:

Model B: The model includes natural mortality only, assumed to follow a normal distribution centered on 0.1 with a variance of 0.04 (Figure 27). No immigration was included, therefore N_1 is the only parameter to be estimated.

Model C: This model includes natural mortality and immigration, which is assumed to occur at a constant rate (number of fish entering per day). Parameters estimated by the model are $\overline{N}$ or N_1 and β.

Model D: To utilize information in the fishery CPUE data, I included CPUE as auxiliary information in model B. I computed average fishery CPUE (sablefish per hook) for each of the 5-day periods based on catch data and assumed that fishery CPUE was proportional to total sablefish abundance:

$$CPUE_i = q * N_i$$

where q is a catchability coefficient, and that $CPUE_i$ followed a normal distribution. To estimate parameters via maximum likelihood, I included the normal likelihood component in the maximization:

$$L(CPUE_i \mid \sigma_i, q, N_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \cdot \exp\left(\frac{(CPUE_i - qN_i)^2}{2\sigma_i^{\,2}} \right)$$

Because the CPUE for each period was averaged across k_i individual CPUE observations (corresponding to individual "trips" or portions of a trip), the variance of the residual for each period was assumed to be inversely proportional to the number of CPUE estimates (k_i) that were available for each period:

$$\mathrm{var}(CPUE_i) = \sigma_i^2 = \sigma^2 / k_i$$

where σ^2 is an overall variance term. To estimate parameters of model D, which now include q and σ^2, I maximized the combined likelihood components by minimizing the sum of log-likelihoods. Both likelihood components were weighted equally (binomial likelihood for mark-recapture data and normal likelihood for CPUE data). The relative weight for the fishery CPUE data could be reduced if the data are deemed unreliable or questionable. In addition, the model was fit in a Bayesian formulation with vague priors on all parameters except m, which, as before, was assumed to follow a normal distribution centered on 0.1 with a variance of 0.04.

Model E: This model extended model C by including fishery CPUE data in the same way that model D extended model B. Parameters estimated by this model are N, β, m, q, and σ.

Extensions of Models D and E that account for differences in size selectivity: I included a correction factor for size selectivity in the Bayesian version of models D and E with a uniform prior distribution between 0.75 and 1, which are likely lower and upper bounds for the selectivity correction factor based on reasonable assumptions about the range of likely selectivities.

OTHER MODELS AND ADJUSTMENTS NOT IMPLEMENTED

Several alternatives to the above models were explored but were not included in the assessment. One alternative for explaining the observed decrease in the ratio of marked to unmarked fish is that marked sablefish experience higher mortalities than unmarked fish. To examine this possibility I fit a model that incorporates differential mortality between marked and unmarked fish. The resulting model provided a much better fit but implies a substantially higher mortality rate of marked fish compared to unmarked fish (approximately 10–20 times higher mortality for marked fish). Such a large mortality of marked fish seems unrealistic, given that in previous years a large number of tags were recovered in subsequent years. The mortality parameter was also difficult to estimate and was highly uncertain, therefore this model was not further developed here. If this model were true and marked fish have a higher mortality, it could have important consequences for estimates of abundance. The model implies that the true number of marks available in the population (M_i) is much lower than estimated by other models that assume equal mortality rates for marked and unmarked fish. Because of the lower estimates of M_i, estimated abundances are substantially lower under this model. This is in stark contrast to the age-structured analyses, which suggest considerably higher abundances.

Several other model runs were completed to freely estimate natural mortality. However, they resulted in highly unrealistic and extremely uncertain estimates for m (ranging from -5 to +5). Therefore, these models were not included here. However, the consequences of uncertainties in m were evaluated by allowing m in the Bayesian analysis to vary from approximately 0.05 to 0.15 and assessing the effect on the resulting estimate, which was relatively minor. The difficulty in estimating natural mortality was not unexpected because the parameter m is notoriously difficult to estimate in stock assessment models. The results suggest that the effect of natural mortality on abundance estimates is relatively minor.

RESULTS

Model Estimates and Goodness of Fit

Abundance estimates were very similar across models because all models had a similar structure and used the same set of mark-recapture data. Results are summarized by model in Table 4. The simple time-stratified estimate (model B) is somewhat higher than the Petersen estimate, in spite of assuming a reduction in the number of marks available to the fishery (by including natural mortality). However, the model appears inadequate because it cannot reproduce the decreasing trend in the ratio of marked to unmarked individuals over time (Figure 4). This is apparent in the residual trend in the number of marks (observed - expected, Figure 29), as well as in comparing the trend in estimated abundance (N_i) to the corresponding Petersen estimates for each period (Figure 30).

Relative to the fishery, the proportion of marks recovered in the longline survey was much lower than expected based on this (and other) models, suggesting either that the number of marks are underestimated in the longline survey or overestimated in the fishery. However, the number of marks is generally highly variable among periods and may have differed due to chance. In previous years, the proportion of marked fish in the longline survey was typically higher or similar to the estimate from the fishery. Since the sampling design did not change in 2006, there is no *a priori* reason to believe that the low number of recovered marks in the longline fishery in 2006 was due to any differences between the portion of the sablefish population sampled by the longline survey and the portion sampled by the fishery.

With immigration included (model C), the fit improves markedly (Figure 31) and suggests a substantial increase in sablefish abundance over time due to immigration of almost 12,000 sablefish per day into Chatham Strait. The improved model fit is evident in the improved likelihood and the reduced value of the model selection criterion AIC (Table 4). The average abundance over the fishing season from model C is very similar to the estimate of initial abundance from model B. However, the estimated immigration rate of almost 12,000 sablefish per day (95% credibility interval: 3,900–20,500) implies an increase in total abundance from 2.18 million sablefish at the beginning of the fishing season to 2.89 million sablefish at the end of the fishing season (Figure 31) and a total immigration of more than 1,200,000 sablefish into Chatham Strait over the 103-day period between Aug. 4 (middle of the longline survey) and November 15 (end of fishery). This appears to be unreasonably high although some immigration into Chatham Strait is believed to occur during summer and fall. Moreover, it is unlikely that immigration occurs at a steady rate over this entire period. Therefore, I do not recommend the use of model C in spite of the improved statistical fit until independent estimates of immigration into Chatham Strait can be examined.

To “anchor” the apparent within-season trend in abundance, models D and E used available CPUE data from the fishery, which display a variable but decreasing trend over time (Figure 28). The decreasing trend, although not significant (weighted linear regression: $t = -1.178$, $p = 0.256$) contrasts with the apparent increase in abundance suggested by model B. Model D (no immigration) fits the CPUE data well (Figure 32) but, like model B, has a clear trend in the mark-recapture residuals (observed minus expected recaptures). The model cannot account for the observed decrease in the number of marks over the season (as evident in an obvious trend and negative residuals in the last part of the fishing season in Figure 33) and displays a trend very different from the corresponding time-stratified Petersen estimates (Figure 34). Model E (with immigration) provides an improved fit to the observed number of marks (Figure 35) and better agreement with time-stratified Petersen estimates (Figure 36), but a worse fit to the catch data compared to model D (Figure 37). The model has a slightly improved total likelihood overall and a lower AIC, thus would be preferred on statistical grounds, although the likelihood ratio test suggests that the added immigration parameter is not significantly different from zero at the 95% level ($p = 0.060$, Table 4). The individual likelihood components confirm that the model provides a better fit to the mark-recapture data, but not as good a fit to the fishery CPUE data (Table 4). The estimated immigration rate is close to 6,000 sablefish per day, implying a total immigration of approximately 600,000 sablefish into Chatham Strait over the course of the fishing season.

Uncertainty

Compared to the Petersen estimator, I introduced additional uncertainty associated with natural mortality and immigration. On the other hand, using a stratified estimator and additional information from the fishery CPUE data could reduce uncertainty about abundance. The estimates of N_1 from models B and D (without immigration) had very similar standard deviations to the Petersen estimator, while the average abundance estimate from models with immigration had somewhat larger standard deviations due to uncertainty associated with the extra parameter β. Allowing the natural mortality rate m to vary randomly between about 0.05 and 0.15 in models

B–D had very little effect on the estimates and standard errors, therefore these estimates are robust to a reasonable range of values for natural mortality.

A potentially larger source of uncertainty is related to selectivity patterns in the fishery and the potential for different selectivities of marked and unmarked sablefish. The most likely effect of the observed selectivity patterns is overestimation of exploitable abundance Appendix A.1. We developed Bayesian versions of models D and E that incorporate effects of fishery selectivity using a correction factor with a uniform prior distribution between 0.75 and 1 (the correction factor reflects the proportion of marked fish available to the fishery). The data appeared to be informative with respect to the correction factor and its posterior distribution suggested that the true correction factor is closer to 1. This agrees with evidence from size composition data and from other models that the selectivity of marked fish is close to 1 for all size classes. The resulting abundance estimates were somewhat lower than those from other models due to a value for the correction factor that was slightly less than 1 on average. As expected, the estimates had considerably wider (and more realistic) credibility intervals:

Model D: N_1 = 2,247,000 (95% CI: 1,891,000 – 2,592,000, sd = 187,900).

estimate of mean correction factor: 0.90 (i.e., 90% average selectivity)

Model E: average N = 2,104,000 (95% CI: 1,742,000 – 2,474,000, sd = 196,500)

estimate of mean correction factor: 0.89, estimate of immigration: 5,560 fish/day

These estimates should be considered tentative because the correction factor does not reflect the apparent difference in selectivity for tagged and untagged fish. Therefore, I consider it premature to use the model for management purposes at this stage, although it would be preferable to uncorrected estimate.

Model Selection and Preliminary Preferred Model

Statistically, models C and E (with immigration) clearly capture the pattern in observed marks over time much better. While the immigration rate in model C may be unrealistically high, the rate estimated by model E seems to be more reasonable, especially considering the high uncertainty (which includes 0). Although the immigration parameter is only marginally significant in model E ($p = 0.060$), the considerably better fit of model E to the mark-recapture data (MR likelihood components: 56 vs. 59), which is likely to be more reliable than the CPUE data, suggests model E as the preferred model. If fishery CPUE data are judged to be unreliable, model B or model A would be preferred (given the high apparent immigration rate in model C).

The Bayesian model with a correction for fishery selectivity is quite promising and is likely to capture the true uncertainty much better. However, the model needs further development before it can be used for management purposes. Nevertheless, because the model suggests that uncertainty in the abundance estimates is at least twice as high when accounting for uncertainty in size selectivity, I used the estimates from model A with twice their standard error for fitting the statistical age-structured assessment model.

Using Data from Multiple Years

I explored a further extension of the mark-recapture estimates of abundance by including estimates over multiple years in a simple time-series model. Consistent methodology (pot survey to mark individuals, recoveries from the fishery) has been used since 2000 and the pot survey design and the quality of catch sampling have improved over time. Assessment authors used the

Petersen estimator to estimate abundances in 2001 and in 2003–2006, while an exploitation rate estimator was used in 2000 and 2002 to convert total catches to total exploitable biomass. I converted these earlier estimates of biomass to numerical estimates of abundance using average weight in the fishery to obtain a time series of numerical abundance from 2000 to 2006 (Figure 38).

The time series of estimated abundance (Figure 38), in spite of relatively narrow confidence intervals, displays a level of variability that may be biologically unrealistic given the life history of sablefish, which are a long-lived species with relatively low natural mortality. Therefore, adult (or exploitable) abundance in a given year is not expected to vary substantially from the previous year. Such slowly changing dynamics are commonly approximated by a first-order autoregressive process, in which the abundance in year *t* fluctuates around a long-term mean and the difference from the long-term mean depends on the difference in year *t-1* through a parameter ϕ plus some random deviation due to estimation errors or other processes:

$$y_t = \phi \cdot y_{t-1} + \varepsilon_t$$

where $y_t = N_t - \overline{N}$ and $\overline{N}$ is the long-term average abundance. In this model, abundance in a given year tends to be "close" to the previous year's abundance (as long as $\phi > 0$), while "reverting" towards the mean. Using 2000–2006 abundances, the model estimated a negative auto-regressive coefficient ($\phi = -0.48$), implying that a year of high abundance is followed by a year of low abundance. The negative coefficient is largely due to relatively low abundance estimates in 2000 and 2002 and a high 2001 abundance estimate. The estimates from the autoregressive model ($\overline{N} + y_t$) fit the time series reasonably well (fitted line in Figure 38), but a negative autocorrelation coefficient is clearly an unrealistic feature of a sablefish abundance time series.

An autoregressive or other appropriate time series model could be used to forecast 2007 sablefish abundance in Chatham Strait (with appropriate confidence intervals). However, because of the unrealistic negative autocorrelation in the short time series, I do not recommend using a time series approach for forecasting future abundances. Moreover, given that a wealth of biological information on natural mortality, age composition, maturity and abundance of pre-recruit stages is available from the longline survey and fishery sampling, the current approach to projecting future abundance and biomass is much more appropriate than a simple time series model. The projection approach does however require a good estimate of current abundance. Because of uncertainty in the mark-recapture estimate for the current year, ADF&G has opted to base projections on the lower 90% confidence bound of projected biomass. This precautionary measure was taken in part to account for unrealistic variability in the estimated abundance or biomass from year to year (e.g., low abundance in 2005). One approach to stabilizing estimates of current abundance would be to smooth the time series of abundance estimates. Because of the negative auto-regressive coefficient, the time series model fit to the 2000–2006 abundances failed to smooth the time series. A simple and straightforward way to smooth the time series, while incorporating multiple years of data into an estimate of current abundance, would be to compute a simple average of the estimates over the most recent years (e.g., 3–5 years). This would produce a more stable estimate of current abundance but could be dangerous practice in the presence of a strong decreasing trend. I do not recommend this approach for Chatham Strait sablefish because there are indications that the population is decreasing and because the ASA model can provide a "smoothed" estimate of abundance over time that is fit to the mark-recapture data.

MARK-RECAPTURE RECOMMENDATIONS

Based on my understanding of the sampling design used for the tagging studies, exploratory analyses of primarily the 2006 tagging data, and my attempts to model the important processes that occur during the marking and recapture phases, I offer the following recommendations:

Data and Pot Survey Design

1) Since the beginning of the tagging program there have been a number of changes in the survey design with respect to the sampling gear used, tags used, size of fish marked, timing of the survey, spatial distribution of sampling effort, and other aspects. Most of these changes were made in response to specific concerns that arose after analyzing tagging data and they have led to a greatly improved tagging program. While a number of aspects of the sampling could be further improved, I strongly recommend that the basic sampling design that has been used for the last several years be kept consistent to establish and maintain a consistent (if imperfect) time series of absolute abundance estimates.

2) Important aspects of the current design that should be maintained include:

    a) Fish are tagged using a pot survey. This appears to have eliminated earlier concerns over gear-shyness, although a full evaluation of this effect is not possible because a large change in the timing of the survey coincided with the gear change. Nevertheless, I recommend continuing the pot survey for marking sablefish in order to maintain continuity until such time that gear shyness of fish tagged by longlining can be evaluated.

    b) Sampling effort in the pot survey is currently distributed throughout Chatham Strait to approximately reflect the distribution of the fishery. This is important to ensure that all tagged fish have approximately the same probability of re-capture, in particular because movement (hence mixing) of sablefish between the time of tagging and the time of the fishery is somewhat limited on the scale of the entire Strait, although it appears sufficient to ensure mixing within sampling strata. Therefore I recommend that the distribution of pot survey effort in proportion to fishery effort continue. However, depending on the year-to-year variability and behavior in the fishery, fishery effort by stratum should be averaged over several years to avoid large interannual changes in the distribution of survey effort. In addition to distributing survey effort by stratum, I recommend that survey effort be apportioned to several depth strata within each geographic stratum to reflect the approximate depth distribution of the fishery because of large differences in size by depth and because movement across depth gradients may be limited.

    c) Double tagging is used to mark sablefish (fin clips + external tags). Although population estimates could be based on fin clips alone, the tags provide valuable information about emigration of sablefish from Chatham Strait and about multi-year trends. Therefore I recommend continuing the use of both fin clips and tags.

3) Considerations regarding the length of fish to mark include:

    a) The pot survey tends to capture smaller fish than the longline survey and the fishery, and tagging in recent years was limited to fish larger than 500 mm fork length. I recommend that all fish exceeding the minimum size captured in the fishery be tagged. Although including smaller fish could potentially bias population estimates if they are not fully vulnerable to the fishing gear (i.e., their probability of capture is smaller than that of larger fish), there are ways to correct for this bias and tagging of smaller fish will help to estimate the vulnerability of all size classes to the longline gear and to the fishery (i.e., their selectivity curves). In contrast, using a larger cut-off value for tagging (e.g., 560 mm, the size above which most fish are believed to be fully vulnerable to the gear and retained in the fishery) results in abundance estimates for an unknown proportion of the vulnerable population, presumably close to 1. Moreover, limiting the estimates to larger size classes misses an important component of the population, i.e., the younger, incoming year classes.

   b) If issues with size selectivity cannot be addressed satisfactorily through modeling, it may be preferable to use fish over 560 mm only in the mark-recapture analysis to obtain conservative estimates of abundance and avoid the interannual variability associated with uncertainties in the selectivity of smaller size classes.

4) Potential improvements that could be made to the pot survey and not disrupt the time series of mark-recapture estimates, but would likely improve abundance estimates and/or provide additional information include:

   a) Increasing the number of large sablefish marked in the survey. Large fish are potentially under-sampled by the pot gear (relatively few large fish are tagged in the pot survey as suggested by length-frequency composition in 2006 – this is at least in part a consequence of more large fish being discarded due to predation by fleas).

   b) Exploring the potential of standardizing effort to obtain another fishery-independent CPUE index. However, given the relatively small survey effort, the index may be of limited utility.

   c) As indicated above, stratifying pot survey effort by depth stratum may further improve the spatial distribution of tag releases.

   d) If logistically possible, conducting the pot survey closer to the time of the fishery to reduce concerns about closure. However, this is a relatively low priority as natural mortality and emigration had a minimal effect on estimates in 2006 and immigration between the pot survey and the longline survey does not affect the models that include immigration (because N_1 estimates the number of sablefish at the time of the longline survey and any immigration prior to the longline survey is not explicitly modeled).

Longline Survey Design

5) Longline stations are sampled following a stratified random sampling design to ensure that the age and size composition in the survey reflects the age-and size composition in the population. The survey design appears sound and should be kept consistent to the extent possible.

6) In my examination of age and size composition and for the age-structured analysis, I assumed that the number of stations in each stratum is proportional to the area of the stratum and that the number of fish measured at each station is proportional to the CPUE at that station. I did not further examine these assumptions. If the assumptions are not met in the current sampling design, each station should be weighted by the area of the corresponding stratum and each specimen measured at a station should be weighted by the CPUE at that station to obtain an unbiased estimate of the overall age/size composition. However, because there were only small differences in size composition between statistical areas in the longline survey, any bias associated with unequal sampling effort in different strata would likely be very minor.

7) The same 44 stations have been sampled by the longline survey since 1997 and I recommend that sampling of these same stations continue in coming years to obtain a consistent CPUE estimate and consistent estimates of age, length, weight, sex, and maturity. If my assumption that the number of samples in a stratum is proportional to stratum CPUE is violated, I recommend using appropriate weights to estimate size/age composition, etc., rather than a change in survey design.

8) In the long run, if the age-structured model is used to estimate abundance, it may not be necessary to maintain two independent estimates (or indices) of abundance (mark-recapture estimate and longline CPUE) on an annual basis, particularly if the longline CPUE data are non-informative (which may be the case), as well as two sources of age-composition data (longline survey and fishery). Instead, resources could be freed up for other priorities.

Fishery Sampling

9) I did not examine the current sampling design but found some evidence in the length composition data that the average size in the total catch based on dividing total catch weight by the total number of fish in the catch is smaller than the average weight of fish in “random” fishery samples, suggesting a potential bias towards larger fish in the sampling process. I suggest reviewing current sampling protocols to ensure that samples are truly random and that this potential source of bias be further examined. If necessary, the sampling protocol may need to be adjusted and/or a correction may need to be applied to the size composition data (for example by estimating a size-dependent “probability of sampling” in the form of a logistic curve).

10) Accurate counts of the number of clipped fish are important for the mark-recapture estimates. I suggest conducting some experiments to assess these counts for potential bias and to determine within- and between-observer variability.

Mark-Recapture Modeling

11) To obtain reliable estimates of absolute abundance from mark-recapture data, I recommend that ADF&G develop and use a general modeling framework to explore different estimators, their sensitivity to assumptions, and potential ways to account for or correct the estimators if assumptions are violated. Such a modeling framework will assist in the choice of the most robust estimators and allows for model extensions and development of new estimators to

relax assumptions if indicated by the data. This will help to establish and maintain a consistent time series of absolute abundance. At present, there do not appear to be any obvious problems that cannot be accounted for in a modeling context and would necessitate a re-thinking of the survey design used to mark and recapture sablefish. This report includes templates for several candidate models that were fit to the 2006 data and could be updated in future years with new data.

12) The recommended candidate models are extensions of the simple Petersen estimator, as outlined in the Extensions of Petersen estimator section, that were developed to address potential concerns over violations of the mark-recapture assumptions to provide improved estimates of population size. The time-stratified estimators that were explored here yield generally similar results to the Petersen estimator, but allow an explicit assessment of some of the assumptions. The use of a time-stratified estimator was prompted by an apparent trend in the ratio of marked to unmarked fish over the course of the fishing season that was apparent in the fishery in 2006 and in several other years. Advantages of the time-stratified models include:

    a) the models allow explicit estimation of immigration rates and an exploration of the effects of emigration and immigration on estimates of abundance

    b) models can account for natural mortality over the course of the prolonged fishing season (the effects of natural mortality are relatively minor, but natural mortality in combination with emigration may have an appreciable effect)

    c) models allow tuning of abundance estimates by time period to fishery CPUE data

    d) models allow for the inclusion of additional sources of uncertainty, for example by incorporating random differences in capture probabilities of marked fish among periods (related to the patchy distribution of fish and/or the behavior of the fishery)

13) In addition to developing a (limited) suite of potential candidate models for the assessment, I strongly recommend that these models be reviewed by a small group of stock assessment scientists, biologists, and managers (as well as by interested fishermen and other stakeholders, if possible) to review a reasonable set of models and their assumptions and to choose a final model for management based on statistical and biological considerations. Such reviews are invaluable in providing an independent check of the models and assumptions used, enhancing transparency, and building trust among different stakeholders.

14) To improve models based on mark-recapture data, I further recommend that other, independent data sources be examined to verify assumptions or parameter values used in the model. For example:

    a) Historical data from NOAA, ADF&G, and DFO tagging programs could be used to obtain independent estimates of the timing and magnitude of immigration into Chatham Strait to confirm or refute the apparent high immigration rates estimated by time-stratified models and to help develop a more realistic representation of immigration in the models.

   b) Analysis of historical Chatham Strait tagging data and tag recoveries from outside Chatham Strait could provide improved estimates of emigration rates that can then be incorporated into mark-recapture estimates of abundance as described in Emigration.
   c) Development of the age-structured model will provide improved estimates of selectivity patterns in the fishery to correct estimators of abundance for size selectivity. Because I constructed the age-structured model after the mark-recapture analyses, estimates of size selectivity for the selectivity correction factor are currently based on NOAA data (Appendix A.1).

15) An important source of bias in the mark-recapture estimates is the discarding of small untagged fish in the fishery, while retaining small marked fish. In this report, I suggest a method for correcting estimates based on differences in size selectivity that needs further development to include effects of differences in selectivity between marked and unmarked fish. However, I consider further development of the size selectivity correction to be a lower priority than developing the age-structured model as the impact on abundance estimates (relative to the basic Chapman estimator or time stratified estimators) was relatively small.

16) Incorporating size selectivity in the models did, however, provide more realistic estimates of the standard errors of abundance for inclusion in the age-structured model. If the current, narrow confidence intervals of the modified Petersen estimator are used in the age-structured model, unrealistic recruitment trends are estimated by the model to try to fit the apparent low abundance in 2005 (as estimated from tagging data).

General Recommendations

17) The current assessment approach has led to a “proliferation” of interconnected spreadsheets with multiple versions of the same data, which may make the analysis more prone to errors or inconsistencies. To simplify updating the assessment annually, I recommend carefully documenting the steps needed for extracting the required data and compiling the “final” data sets that are used in the analysis in a single spreadsheet for easy reference.

AGE-STRUCTURED ASSESSMENT

HISTORY OF AGE STRUCTURED ASSESSMENT IN CHATHAM STRAIT

The first age-structured model for Chatham Strait sablefish was developed in 1995 based on 1989–1995 data. The model was implemented using a spreadsheet version adapted from an assessment model developed for Gulf of Alaska sablefish by Sigler (1999). The model performed poorly, presumably because of the short time series of available age data, and was felt to result in unrealistically high biomass estimates. An updated ASA in 1998 produced better fits to the data and more plausible results, but was ultimately not used for stock assessment purposes because resulting biomass estimates were still considered to be unrealistically high. The ASA was last updated with 1988–2001 data. Although results seemed reasonable, mark-recapture estimates of abundance were continued to be used to set quotas and the age-structured assessment has not been updated since that time.

I developed a statistical age-structured model for the Chatham Strait sablefish population by first updating and modifying a previously developed spreadsheet version of an age-structured assessment (Dressel, personal communication[4]). The model was updated with data through 2006. To improve the efficiency and speed of estimation and to obtain estimates of uncertainty for quantities of interest, I implemented the model in ADMB, automatic differentiation software that was developed as a set of libraries under the C++ language. ADMB is currently the software of choice of west coast stock assessment scientists for fitting non-linear models with a large number of parameters.

The model implementation was based on an ADMB template obtained from the Alaska Fisheries Science Center (Dana Hanselman[5], personal communication.). The ADMB template was originally developed for yelloweye rockfish and was substantially modified to adapt it to the specifics of the Chatham Strait sablefish population and fishery. This section describes the basic model structure and presents results for several alternative models based on data through 2006.

CHANGES TO ASA FROM 2001 VERSION

The following changes were made to the previous version of the model:

1. I estimated separate fishery and survey selectivity curves to account for apparent differences due to sorting (selecting against small fish) in the fishery. These were previously assumed to be the same.
2. I fixed several issues with the previous ASA:
    a. "Available number of fish" in the longline survey was estimated by essentially (and incorrectly) subtracting the catches, i.e., it was assumed that the fishery occurred prior to the survey, which has not been the case since 1988.
    b. Annual fishing mortality rates are estimated as free (or constrained) parameters in the new ASA. In the previous model, exploitation rates (for fully exploited age classes) were estimated after the fact by converting estimated catches (weight) to numbers using estimates of mean weight, then dividing by the estimated number of available fish.

    Because of these issues, and to make the model more compatible with recent approaches to statistical age-structured assessments (to facilitate implementation in ADMB using the yelloweye rockfish ADMB template), I modified the model by estimating annual fishing mortalities and assuming separability (i.e., $F_{t,a} = s_a * F_t$). The model is described in detail below.
3. I changed the model to estimate survey catchability, q, rather than using a fixed value (as was done in the previous version).
4. I used two separate time series of CPUE based on 1 hour soak time (1988–1996) and based on 3+ hour soak time (1997–2006) with separately estimated catchability

---

[4] Dressel, S.C., Alaska Department of Fish and Game, Region I, Division of Commercial Fisheries, P.O. Box 110024, Juneau, Alaska 99811-0024.

[5] Dana Hanselman, Alaska Fisheries Science Center, NOAA, 17109 Point Lena Loop Rd, Juneau, AK 99801

coefficients. In the ADMB implementation, q can be constrained by specifying appropriate prior distributions, but uninformative priors are used by default.

5. The previous model was tuned to fishery CPUE as an index of abundance, but did not include a component to fit the catch series (or catch-at-age matrix). Therefore, I added a likelihood component for fitting the catch data (total catch weight by year).
6. The previous model included age classes 2–19 and an age 20+ group. Because the number of fish aged by age group was relatively high through at least age 25, I extended the age range to include age groups 2–25 and an age 26+ group.
7. The previous model fit both size composition data and age composition data from the longline survey. Because this essentially uses the same length data twice, I used either age compositions (where available) OR size compositions (when only lengths were available: 2000 and 2001 fishery data).
8. The model was modified to estimate $F_{40\%}$ and the corresponding Allowable Biological Quota (ABC) internally, rather than using external estimates of $F_{40\%}$ to calculate the quota.
9. Separate weight-at-age estimates from the fishery and longline survey data were used to convert numbers to biomass to (a) fit the catch data series and (b) compute total, spawning, and exploitable biomass within the model as explained below.
10. An aging error matrix was used to convert the “true” (modeled) age compositions to estimated age compositions for the longline and fishery samples.
11. Other, minor changes were introduced in the ADMB implementation of the model as described in detail below.

INPUT DATA

To fit the age-structured model, the following data sources, parameters, and settings were used:

- Fishery catches from 1980–2006 were used in model fitting and were assumed to be accurately and precisely estimated (large weight on sum of squared catch deviations).
- Two separate survey indices were used as independent indices of abundance (the longline survey was assumed to occur August 1 for computing predicted CPUE):
    - Longline survey CPUE, 1988–1996 (1-hour soak time).
    - Longline survey CPUE, 1997–2006 (3+ hour soak time).
- Fishery CPUE data from 1980–2006 was used as a fishery-dependent index of abundance.
- Mark-recapture estimates of abundance for 2003–2006 were used as either an absolute or a relative index of abundance in model fitting, depending on the model run.
- Longline survey age compositions for ages 2–25 and a plus group (26+) for older fish.
- Fishery age compositions for ages 2–25 and a plus group (26+) for older fish.

- Fishery size compositions for 2000–2001 (length classes 465–1065 mm in 20 mm increments).
- An age-length transition matrix to convert modeled numbers at age to numbers at size was estimated from the fishery survey data for 2002/2003, the two years closest to 2000–2001 that had both age and size data.
- Weight-at-age was assumed to be constant over time and was estimated from random samples obtained during the fishery and longline surveys.
    - Average weight-at-age (males & females combined) of fish sampled from the 2002–2006 fisheries was used to compute predicted catches.
    - Average weight-at-age (males & females combined) of fish caught during the longline survey (1988–2006) was used to compute total biomass and to estimate ABC.
    - Weight-at-age of female fish caught during the longline survey (1988–2006) was used to compute spawning biomass.
- The proportion of female fish that were mature at a given age was estimated from longline survey data for 2000–2006 to compute spawning biomass.
- The proportion of female fish in the longline survey was estimated by age (1988–2006 data) to compute spawning biomass.
- An ageing error matrix from the NMFS sablefish assessment was used to allow for imprecise age readings.
- Spawning was assumed to occur on February 1 (for computing spawning biomass).
- Natural mortality was assumed to be 0.1 or was estimated with a relatively tight prior with a mean of 0.1.

Fishery Catches

Total catches (biomass) for 1980–2001 were taken from a previous ASA model and were updated with recent catches for 2002–2006. I divided round weight (pounds) by 2.20462 to convert weight to kilograms, assuming that biomass for 1980–2001 are also expressed as round weight.

Abundance Indices

The survey CPUE was treated as 2 separate time series:

- 1988–1996: CPUE based on 1 hr soak time
- 1997–2006: CPUE based on 3+ hour soak time

Standardized CPUE data were not available for the entire time period. If an improved series of CPUE becomes available it should be substituted for this series. However, slight changes to the time series are likely to have a negligible effect on results as the model fit does not appear to be strongly affected by the inclusion or exclusion of the CPUE series.

Mark-recapture estimates of abundance for 2003 to 2006 (Figure 38) were used either as an absolute measure of abundance ("catchability", *q*, assumed to be 1), or as a relative abundance index, similar to longline and fishery CPUE. When used as an absolute measure of abundance, it was assumed that mark-recapture estimates provide an accurate estimate of sablefish abundance at the time of the longline survey.

Age Composition Data

Age data are available from the longline survey for 1988–2006 and for the fishery from 2002–2006. The number of fish aged by year varied from 206 to 831 for the longline survey:

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
N	307	222	206	274	209	431	335	246	313	577	357	296	410	720	669	790	765	831	741

Sample sizes for the fishery were considerably higher, but ages are only available since 2002:

Year	2002	2003	2004	2005	2006
N	2325	1801	1804	1404	1684

The observed proportion of fish by age and year was estimated from these data. Because of the relatively small sample sizes in the longline survey there were many zero values in the resulting year-by-age matrix, which could lead to numerical problems when computing the likelihood, hence small constants were added to individual likelihood components.

In the earlier ASA model, age composition from the longline survey was used to represent age composition in the fishery, and fishery and survey selectivity were assumed to be the same. I used separate selectivity schedules to compute expected age compositions to fit to the observed age composition data for the longline survey and for the fishery. There were some differences in age compositions between the data in the 2001 ASA and the age compositions in more recent records. I was unable to resolve these differences and used the most recent version of 1988–2006 data.

The age composition data show evidence of strong year classes moving through the population (Figure 39, Figure 40) which is essential for the age-structured analysis to provide reasonable estimates. Age compositions for the fishery and for the survey were computed using R.

Age-Length Transition Matrix and Observed Length Composition

Fishery size compositions, rather than age compositions, were fit in the model for the years 2000 and 2001 (Figure 41) because no age information was available for the fisheries in these years.

Observed size compositions were estimated from random fishery samples. The number of length samples varied widely and ranged from a minimum of 212 to a maximum of 6,799:

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
N	1606	248	222	288	212	3855	3111	6799	2891	5832	5144	2811	418	744	671	809	767	847	1482

The proportion of fish in 200 mm size bins was estimated from these data for each year, but only the size compositions for 2000 and 2001 were used in the ASA model.

To get the predicted length composition for 2000 and 2001, I constructed an age-length transition matrix based on the combined data for 2002 and 2003, which were the closest years to 2000/01 with both age and size data, and had very similar length-at-age patterns. I fit a Ludwig von Bertalanffy (LVB) growth model to the 2002/03 data and constructed the transition matrix from fitted model values for length-at-age (L_a) and from the estimated distributions of length at each age:

$$L_a = L_\infty\left[1 - e^{-k(a-a_0)}\right]$$

where L_∞ is the estimated maximum length, k is a growth parameter and a_0 is the x-axis intercept. The fitted model parameters are shown below and the fit is depicted with 2 standard deviations indicated at each age in Figure 42. The variance was assumed to be constant across ages (Figure 43), except for the 26+ group, although there was some evidence for heteroscedasticity.

	Estimate	Std Error	*t* value	*P*-value
k	0.061	0.011	5.492	<0.001
L_∞	800.862	21.117	37.926	<0.001
a_0	-18.840	3.122	-6.034	<0.001

Predicted proportions in each 200 mm length class for each age were obtained from the density of a normal distribution with mean equal to the predicted average length and standard deviation equal to the estimated standard deviation at age. The proportion of fish in the age 26+ group was estimated separately based on the pooled data using a kernel density estimator, rather than the normal distribution, because the distribution was strongly right-skewed due to the inclusion of many older fish (Figure 44).

The resulting age-length transition matrix was used in the spreadsheet ASA model and in the ADMB model to compute the predicted length composition from the predicted age composition:

$$\hat{\mathrm{p}}_{\mathrm{t},l}^{size} = \hat{\mathrm{p}}_{\mathrm{t,a}}^{age} \times \mathrm{A}_{\mathrm{a},l}$$

where the ***p*** matrices are the predicted size composition (proportion of fish in length-class *l* in year *t*) and the predicted age composition (proportion of age *a* fish in year *t*), and $\boldsymbol{A_{a,l}}$ is the age-length transition matrix (proportion of fish in length-class *l* at age *a*).

Longline Survey Weight at Age

I calculated and examined patterns in weight at age (as well as length-weight, length at age, and maturity at age) based on longline survey data (1988–2006). Age readings have been corrected and updated since these data were compiled, but the re-released ages for the longline survey cannot currently be attributed to the original specimens because there are duplicate specimen numbers for each year and trip. Therefore I used the original ages in the analysis. These can be replaced when the re-released age data is reconciled. For the analysis I selected all randomly sampled sablefish (rows with sample code 1 and species code 710) to examine length, weight, age, sex, and maturity data. I did not exclude any of the ages, regardless of the age readability code, whereas the original analysis apparently used only ages with readability code 1, 2, or 3.

I estimated mean weight at age based on the LVB growth equation for weight. These values are needed to compute biomass at the time of the survey. The observed weights at age (Figure 45) show the large difference between males and females. However, to predict biomass in the ASA I used a growth model fit to the combined data (both sexes, all years). The LVB growth model for weight, which combines length-weight and length-age models, can be written as (e.g., Quinn & Deriso 1999, p.140):

$$w_a = w_\infty \left[1 - e^{-k(a-a_0)}\right]^b$$

where w_∞, k, a_0, and b are the parameters of the model to be estimated. The advantage of fitting the growth model for weight is that ages can be converted to weights in a single step, rather than first estimating lengths from ages and then weights from lengths. The disadvantage is that the 4-parameter model is more difficult to fit. Because the variance of observed weights at age was much larger at older ages, I used a weighted regression with weights inverse proportional to the variance.

The estimated coefficients and their standard errors were as follows:

Parameters:

	Estimate	Std Error	*t* value	*P*-value
k	0.093	0.029	3.147	0.002
w_∞	4.875	0.180	27.022	<0.001
a_0	-4.185	7.930	-0.528	0.598
b	1.451	1.843	0.787	0.431

Fitted mean weights at age for all years combined are shown in Figure 45.

Fishery Weight at Age

Fishery data are available for 2000–2006, but no fish were aged in 2000 or 2001. I used only randomly selected data from the regular fishing season for 2002–2006. I estimated mean weights at age for both sexes combined and for males and females separately using a weighted least-squares regression as for the longline data above. Results are similar but suggest some systematic differences in size between the longline survey and the fishery (Figure 46). Combined weights were used to compute catches and exploitable biomass.

The estimated coefficients and their standard errors were as follows (b fixed at 3 because of convergence problems):

	Estimate	Std Error	*t* value	*P*-value
k	0.091	0.006	15.02	<0.001
w_∞	5.287	0.073	71.94	<0.001
a_0	-13.357	1.155	-11.57	<0.001

Results suggest that at the time of the fishery, average weight at age was apparently between 150g (age 19) and 450g (age 3) heavier than at the time of the longline survey (Figure 46). This may be due to a combination of two factors: (1) growth between the time of the survey and the time of the fishery, even though they are only separated by a few weeks, and (2) size selectivity in the fishery, which may discard “skinny” fish, thus increasing the average weight-at-age of the retained catch.

Similar to weight at age, estimates of LVB growth curves for length at age suggest substantial differences in length at the youngest ages (Figure 47). For example, mean lengths for the youngest age classes are up to 50 mm shorter at the time of the longline survey.

Variability in Length Over Time

One of the concerns in age-structured models is that changes over time in growth, length at age, or weight at age may bias results, which could lead to biased estimates if trends are not accounted for. This happened, for example, in the Pacific halibut stock assessment (Clark et al. 1999). Therefore I looked at changes in length at age by year for female (Figure 48) and male (Figure 49) sablefish in the fishery. Estimated growth differed (significantly) among years with a tendency for increased length at age for younger age classes over time (2002–2006, Figure 50). Data from the longline survey going back to 1988 shows that growth was relatively constant from 1988 to 1998, with reduced growth (smaller *k*, i.e., smaller size at younger ages) in the period 1999–2003/04, and a “recovery” to normal growth in 2005/06 (Figure 51). The L_∞ parameter showed roughly the opposite pattern (parameters are negatively correlated) but is more variable than *k*.

The differences, while statistically significant, are relatively modest. I did not examine changes in weight at age (which is complicated by the necessity for a 4-parameter model or a 2-stage approach), but similar differences in weight at age are likely to exist. I did not incorporate changes in weight at age into the model at this point and I recommend that some sensitivity analyses are done to examine the impact of potential changes in weight at age on model results.

Maturity at Age

Maturity at age data are available from both the longline survey and the fishery. Because the fishery is more selective than the longline survey (i.e., chooses larger fish, regardless of age), both the weight and the proportion of fish mature at a given age are likely to be higher in fishery samples, suggesting earlier maturation than is actually the case. Therefore, I used the longline survey data to estimate a more conservative maturity-at-age schedule. Because maturity at age was anomalously low in 1999, but has been relatively consistent since 2000, a maturity-at-age schedule was estimated from the 2000–2006 longline survey data by fitting a logistic regression to the maturity data. The estimated maturity at age schedule suggests that in recent years 50% maturity was reached at age 6.88 with the following estimated proportions mature at age:

Age:	2	3	4	5	6	7	8	9	10	11	12	13	14
Prop:	7.22	11.61	18.14	27.21	38.67	51.55	64.22	75.18	83.67	89.61	93.57	96.08	97.64
Age:	15	16	17	18	19	20	21	22	23	24	25	26	
Prop:	98.59	99.16	99.50	99.70	99.82	99.90	99.94	99.96	99.98	99.99	99.99	100.00	

Sex Ratio

The longline survey data showed clear evidence that the proportion of female sablefish in Chatham Strait decreases with age while the sex ratio at age stayed relatively constant over time I therefore pooled samples from all years and estimated the proportion of females by age class using a logistic regression (Figure 52):

Age:	2	3	4	5	6	7	8	9	10	11	12	13	14
Prop:	54.4	54.7	54.9	55.0	55.1	55.2	55.1	54.9	55.6	54.1	53.4	52.5	51.5
Age:	15	16	17	18	19	20	21	22	23	24	25	26	
Prop:	50.3	49.2	48.0	46.9	45.8	44.7	43.6	42.6	41.6	40.6	39.8	34.2	

These proportions are used in the age-structured model to estimate spawning biomass.

Aging Errors (Age-Age Transition Matrix)

To allow for uncertainty in the aging process, I used the NOAA age-age transition matrix that is currently used for sablefish in the Gulf of Alaska/Bering Sea. Even though ages as read by NOAA and ADF&G may differ, the relative aging error is likely to be very similar between readers. If enough data are available from the ADF&G age reader and if aging may be different, a new age-age transition matrix could be developed in the future. The age-age transition matrix is used to convert the "true" age composition estimated by the population dynamics model to an "expected" age composition that reflects age reading errors (i.e., the number of fish at a given age is spread out over several age classes to account for aging errors).

Natural Mortality

Natural mortality was fixed in most model runs and was assumed to be 0.1. This is the value that was used in the previous ASA for Chatham Strait and is the same value used by NOAA in the assessment of the Gulf of Alaska/Bering Sea/Aleutian Island stock of sablefish (Hanselman et al. 2007). Several model runs relaxed the prior on *m* in an attempt to estimate natural mortality. The estimated value generally fell near the lower end of the prior distribution and was unrealistically small when the prior distribution was wide.

PARAMETERS ESTIMATED

A maximum of up to 88 parameters were estimated in the ASA, consisting of:

- 2 selectivity parameters for the fishery (steepness, age at 50% selectivity)
- 2 selectivity parameters for the longline survey (steepness, age at 50% selectivity)
- Average fishing mortality and fishing mortality deviations for each year, 1980–2006 (28 parameters)
- Average recruitment and recruitment deviations for each year (51 parameters), representing:
    - 27 years of age-2 recruitment
    - 24 initial year abundances ages 3–26+
- 4 catchability coefficients, *q*
    - Longline survey, 1988–1996 (1-hour soak)
    - Longline survey, 1997–2006 (3+ hours soak)
    - Fishery CPUE
    - Mark-recapture estimates of abundance (most model runs had a fixed $q = 1$)
- log of natural mortality (if estimated)

Independently (i.e., outside the model) estimated parameters include natural mortality ($M = 0.1$, if not estimated in the model), four LVB growth parameters for each weight-at-age relationship (males, females, or combined sablefish in the longline survey and in the fishery), three LVB parameters for length at age in the fishery to construct the age-length transition matrix, two maturity-at-age parameters for the female maturity schedule (slope, age at 50% maturity), and two parameters for a logistic regression of the proportion of females by age.

MODEL STRUCTURE

The basic model structure is identical in the spreadsheet implementation of the ASA and in the ADMB implementation. However, the ADMB implementation uses priors (informative or uninformative) on a number of quantities and has options for adding a spawner-recruit function in the model to estimate mean recruitment at a given level of spawner abundance. Other, relatively minor differences between the spreadsheet version and the ADMB version are noted below. What follows is a general description of the model structure, for details of the implementations, refer to the description of the spreadsheet and to the detailed description of the ADMB version of the ASA in Appendix B.1.

Population Dynamics

The number of sablefish by year and age ($N_{t,a}$) are computed from basic population dynamics equations:

$$N_{t+1,a+1} = N_{t,a} \cdot S_{t,a} = N_{t,a} \cdot e^{-(F_{t,a}+M)}$$

where $F_{t,a}$ is fishing mortality in year t at age a as determined by annual fishing morality rates F_t and selectivity-at-age s_a (i.e., “separability”: $F_{t,a} = F_t * s_a$), $S_{t,a}$ is the proportion of age a fish surviving from year t to year $t+1$, and M is natural mortality. Numbers in the first year (1980) and recruitment at age 2 in all subsequent years are essentially estimated as free parameters. The ADMB implementation includes a constraint on recruitment variability but this has little impact on overall results. Numbers for the oldest age group (A, which is a plus group: A+) are computed as the sum of the corresponding cohort’s survivors plus the surviving members of last year's A+ group:

$$N_{t+1,A} = N_{t,A-1} \cdot S_{t,A-1} + N_{t,A} \cdot S_{t,A}$$

Catches (by number) at age in a given year are computed from the Baranof catch equation (Eq. 8.25 in Quinn and Deriso 1999):

$$\hat{C}_{t,a} = N_{t,a} \frac{F_{t,a}}{Z_{t,a}} \left(1 - e^{-Z_{t,a}}\right) = N_{t,a} \frac{F_{t,a}}{Z_{t,a}} \left(1 - S_{t,a}\right)$$

where $Z_{t,a} = F_{t,a} + M$. Total expected catch (weight) in a given year is computed by multiplying age-specific catches (numbers) by the average weight at age and summing over age classes:

$$\hat{C}_t = \sum_{a=r}^{A} C_{t,a} \cdot w_a^{fshy}$$

where w_a^{fshy} is the average weight of age-a fish at the time of the fishery (as estimated from 2002–2006 fishery samples), and $r = 2$ is the recruitment age.

Total predicted biomass and (female) spawning biomass for each year are calculated similarly to the predicted catches, but I used weights as estimated from the survey data:

$$\hat{B}_t = \sum_{a=r}^{A} N_{t,a} \cdot w_a$$

$$\hat{SB}_t = 0.5 \sum_{a=r}^{A} N_{t,a} \cdot w_a^f \cdot p_a \cdot r_a$$

where w_a is the average weight at age of sablefish from the longline survey (males and females), w^f_a is the weight at age of female sablefish, p_a is the proportion of females mature at age a and r_a is the proportion of female fish in the population at age a.

Survey CPUE

Survey CPUEs are assumed to be proportional to abundance:

$$CP\hat{U}E_t = q_{srv} \sum_{a=r}^{A} N_{t,a} \cdot s_a^{srv}$$

where q_{srv} is the estimated survey catchability and s_a is the estimated survey selectivity at age and a likelihood component is added for each survey to minimize deviations between predicted and observed CPUE values. The estimated abundances from the tagging experiments in 2003–2006 are treated equivalently using a catchability coefficient (q_{MR}), but the corresponding catchability was fixed at $q_{MR} = 1$ in most model runs, thus mark-recapture estimates are assumed to accurately estimate the abundance of sablefish in Chatham Strait.

Age and Size Compositions

The predicted fishery age composition in year *t* (proportion of total population that is estimated to be in age class *a'* in year *t*, for years with age data) is computed by dividing predicted age-specific catches by the total predicted catch in a year and multiplying by an age-age transition matrix to account for errors in age readings:

$$p_{t,a'} = \sum_{a=r}^{A} \mathrm{M}_{a,a'} \left(\hat{C}_{t,a} \Big/ \sum_{a=r}^{A} \hat{C}_{t,a} \right)$$

where the latter term (ratio) is the proportion of the predicted catch by age class and $\mathbf{M}_{a,a'}$ corresponds to column *a* of a symmetrical age-age transition matrix. The fractions in $\mathbf{M}_{a,a'}$ are the proportion of (true) age-*a* fish that are determined to be of age *a'* by the aging lab.

The predicted survey age composition is computed equivalently:

$$p_{t,a'} = \sum_{a=r}^{A} \mathrm{M}_{a,a'} \left(s_a \cdot N_{t,a} \Big/ \sum_{a=r}^{A} (s_a \cdot N_{t,a}) \right)$$

Fishery and survey size compositions are computed equivalently using the age-length transition matrix and substituting size classes *k* and *k'* for age classes *a* and *a'* in the above formulas to compute the proportion of the population in year *t* that are in size class *k'*.

Selectivity

The model estimates two separate selectivity functions for the longline survey and for the fishery. Both selectivity functions are assumed to follow a 2-parameter logistic model:

$$s_a = \left(1 + e^{-\delta(a - a_{50\%})}\right)$$

where δ and $a_{50\%}$ are the parameters for the slope and age at 50% selectivity, respectively.

Likelihood Components and Objective Function

The total likelihood to be maximized includes the following components. The likelihood is maximized by minimizing an objective function consisting of the sum of negative log-likelihood components, hence all likelihood terms are expressed on the negative log-scale.

Sum of Squared Catches

To fit the catch data, the sum of squared deviations between observed and predicted catches (1980–2006) on the log-scale is minimized (assuming a log-normal distribution of catches):

$$SSQ(catch) = \sum_{t=1}^{n} \left(\log(C_t) - \log(\hat{C}_t)\right)^2$$

(This is simply a sum of squared differences and not a true likelihood because any uncertainty in catches is not accounted for. A large weight is put on the sum of squared catches to fit catch data exactly, assuming that catches are accurately estimated). If the variance of annual estimates of catch were known, it would be easy to modify the model to account for uncertainty in catches by using the same likelihood formulation as for CPUE below.

Survey/Fishery CPUE

Likelihoods for each "survey" (2 longline surveys, fishery CPUE) are included to minimize deviations between observed and predicted CPUE (assuming a normal distribution). The total (negative) log-likelihood is computed by summing negative log-likelihoods over *n* survey years (note that all constants are dropped from the likelihood):

$$-\log L = \sum_{i=1}^{n} \frac{(CPUE_i - CP\hat{U}E_i)^2}{2s_{CPUE_i}^2}$$

where s_{CPUEi} is the standard error of the survey biomass for a given survey.

Mark-Recapture Abundance Estimates

Predicted abundances from the model are fit to independently estimated abundances from recent tagging experiments by minimizing the negative log-likelihoods over the *n* years that have abundance estimates (2003–2006):

$$-\log L(MR) = \sum_{t=1}^{n} \frac{(N_t - \hat{N}_t)^2}{2s_{N_t}^2}$$

where N_t is the mark-recapture estimate of abundance in year t, $\hat{N}_t$ is the predicted abundance in year t, and s_{N_t} is the standard error of the estimated abundance for year t.

Age and Size Compositions

As in most age-structured assessments, I assumed a multinomial distribution for survey age compositions (1988–2006):

$$L_{survey\,age}(y \mid p) = \prod_{t=1}^{T} \prod_{a=r}^{A} \pi_{t,a}^{p_{t,a} n_t}$$

$$-\log L_{survey\,age} = -\sum_{t=1}^{T} n_t \sum_{a=r}^{A} p_{t,a} \log(\pi_{t,a})$$

where $p_{t,a}$ and $\pi_{t,a}$ are the observed and expected age composition (proportion of fish in age class a in year t), respectively, and n_t is a measure of sample size for each year (which is taken to be the square root of the number of fish aged, following Thompson et al. 2006). A small positive constant (0.0001) was added to the observed and estimated proportions to prevent potential problems with zero values in the estimation.

The multinomial likelihood for fishery age compositions, $L_{fshy\,age}$, was computed equivalently for all years with age data (2002–2006). Again, for the multinomial sample size I used the square root of the actual number of lengths sampled.

Assuming a multinomial distribution for lengths by size class, the multinomial likelihood for fishery size compositions (2000, 2001) is computed similarly as:

$$L_{fshy\,size} = \prod_{t=1}^{T} \prod_{l=1}^{k} \pi_{t,l}^{p_{t,l} n_t}$$

$$-\log L_{fshy\,size} = -\sum_{t=1}^{T} n_t \sum_{l=1}^{k} p_{t,l} \log(\pi_{t,l})$$

where $p_{t,l}$ and $\pi_{t,l}$ are the observed and expected size compositions (proportion of fish in length class l in year t), respectively, and n_t is a measure of sample size for each year (which is taken to be the square root of the number of length taken, following Thompson et al. 2006). As for the age compositions, a small positive constant (0.0001) was added to the observed and estimated proportions to prevent potential problems with zero values in the estimation.

Penalty Terms and Priors

A penalty function for recruitment deviations (if a prior is specified for recruitment variability) assumes that recruitments follow a log-normal distribution:

$$-\log L = n \cdot \log \sigma_R + \sum_{i=1}^{n} \frac{(\log R_i - \overline{\log R})^2}{2\sigma_R^2}$$

Currently, recruitment is simply constrained by minimizing squared recruitment deviations, but with a small weight. This seemed to work better and was easier to adjust to get reasonable results.

A penalty term is added for fishing mortality deviations to limit the difference between annual $\log(F_i)$ values (deviations) and the overall average $\log(F)$ across years:

$$penalty = \sum_{i=1}^{n} (\log F_i - \overline{\log F})^2$$

Prior probabilities are added to the likelihood to constrain parameters for which a prior distribution is specified. Prior distributions for the recruitment variability (σ_R^2, if specified) and for survey catchabilities (q_{srv}) are assumed to follow a normal distribution on the log scale with specified means and variances. In general terms, the prior for a parameter θ is calculated as:

$$\pi(\theta) = \frac{\left(\log(\theta) - \log(\mu_\theta)\right)^2}{2 \cdot s_{\log\theta}^2}$$

A very large penalty term is included for SPR calculations to solve for $F_{40\%}$, $F_{35\%}$, etc. This term forces $SSB_{40\%}$ to be equal to (very close to) 0.4*SSB under no fishing, etc, to estimate the corresponding $F_{40\%}$ rate but including this term in the likelihood does not affect any of the other likelihood terms.

QUANTIFYING UNCERTAINTY

For all parameters of interest, asymptotic standard errors and correlations are computed from the Hessian matrix and are returned with ADMB output if requested. In addition, likelihood profiles can be computed for specific parameters of interest using the an option in ADMB.

RESULTS

I fit a number of alternative models that appeared to be reasonable representations of sablefish dynamics in Chatham Strait. Here I review results from a series of ADMB model fits that examine the sensitivity of model results to different assumptions.

In general, the estimated trend in sablefish abundance and biomass over time was very robust, whereas the absolute abundance or biomass estimates are highly variable, which is not unusual in stock assessments and is often the reason for fixing q for some fishery-independent biomass or abundance estimate. Therefore, I fixed $q = 1$ for the mark-recapture estimates of abundance in most model runs, assuming that these provide unbiased estimates of sablefish abundance in Chatham Strait.

One important result was that in all of the scenarios, including those which did not fix q_{MR} at 1, the estimated level of current spawning biomass was less than, usually much less than, 40% of the “virgin” biomass ($B_{40\%}$, which may be considered a proxy for the spawning biomass at which maximum sustainable yield is achieved). At the observed levels of biomass, the harvest control rule used by NPFMC (Figure 1) would require a substantial reduction in fishing mortality from

$F_{40\%}$. I therefore recommend a more precautionary harvest rate than $F_{40\%}$ until the stock is rebuilt to levels closer to $B_{40\%}$.

Because I did not do a comprehensive model evaluation, I show results for the base model only, which fixes natural mortality at $M = 0.1$ and catchability for the mark-recapture estimates at $q = 1$ with uninformative (or no) priors on other parameters.

Abundance/biomass trends (Figure 53): This and all models suggest an increase in biomass in the early 1980s, resulting from several strong year classes in the late 1970s and early 1980s, followed by a long-term decline since 1986.

Recruitment trends (Figure 54): Patterns in recruitment are quite variable, but in this as in all model runs there appears to be a period of high recruitment in the late 1970s/early 1980s and another period of relatively good recruitment in the late 1990s. All models estimated very low recruitment in recent years since about 2001. Therefore biomass is projected to decline further, which is another reason for taking a very precautionary approach in order to maintain an adequate level of spawning biomass.

The model provides a reasonable fit to the fishery CPUE series from 1980–2006 (Figure 55), although the model estimates suggest a continuing decline in recent years, while fishery CPUE has increased or remained stable since the late 1990s.

The longline CPUE data are quite variable and the model does not fit the apparent trend in longline CPUE very well (Figure 56). Interannual variability in the longline CPUE is very high and it appears that the CPUE index is not very informative.

Model estimates of abundance provide a reasonable fit to the mark-recapture estimates, but cannot account for the apparent decline in 2005 and the subsequent increase in 2006 (Figure 57). The model "tries" to accommodate the apparent increase in abundance between 2005 and 2006 by estimating a large recruitment of age-2 fish in 2006, but there is no evidence for a large year class in the age composition data from either the longline survey (Figure 58) or the fishery (Figure 59).

Estimated selectivity schedules for the survey and fishery are quite reasonable, but are highly constrained by the 2-parameter logistic model and by restricting fishery selectivity at age to be less than or equal to survey selectivity (Figure 60).

The estimated exploitation rates (catch/exploitable biomass) suggest increasing exploitation rates from the late 1980s to the late 1990s and a subsequent decrease to moderate exploitation rates around 10% since 2000 (Figure 61). These estimates reflect both changes in exploitable biomass and in management.

The estimate of current spawning stock biomass in relation to the unfished stock size (=depletion) by different models ranged from 13.7% to 17.8% for models with $M = 0.1$ and $q = 1$ for the mark-recapture index. For models that relaxed the constraint on M, q, or both, depletion estimates ranged from 16.8% to 28.2%. Thus, even under the most optimistic assessment model, projected spawning biomass in 2007 is well below $B_{35\%}$ (often considered a proxy for B_{MSY}) and in many cases it is below ½ B_{MSY}, which is the minimum stock-size threshold for federally managed species and calls for a rebuilding plan under the Magnuson-Stevens Act. Under the harvest control rule used by the NPFMC, fishing mortality is substantially reduced when spawning stock size drops below $F_{40\%}$.

SENSITIVITY ANALYSES

I examined the sensitivity of model results to various constraints and to the inclusion or exclusion of some data sets when fitting the model.

Sensitivity to Constraints On Recruitment

Results were not very sensitive to constraining recruitment and estimated parameters only changed appreciably if a large penalty term was used (large value for ‘wt_rec_var’). If no penalty was used (weight 0), recruitments tended to be extremely variable, resulting in a number of years with zero recruitment (Figure 62). With zero or very small weights (0.05), the model was also more difficult to fit and the Hessian matrix was not positive definite. Therefore, I used a relatively small weight in other model runs (wt_rec_var = 0.1), which ensured that recruitment was larger than 0 in all years, but had minimal impact on estimated parameters.

Sensitivity To Constraints On Fishing Mortality:

Similar to recruitment, the effect of changing the weight on squared fishing mortality deviations was relatively minor and over a broad range of weights the pattern of fishing mortality did not change appreciably. Because the patterns in fishing mortality rates (or exploitation rates) seemed quite reasonable, I did not constrain fishing mortality deviations at all in other model runs (wt_fmort_reg = 0).

Sensitivity to Including/Excluding Various Data Sources:

I re-fit the model using various combinations of data sources and compared fits to the “base” model that included all of the data components described above and used penalty weights of 0.1 and 0 for recruitment deviations and fishing mortality deviations, respectively. Excluding the longline survey data resulted in a reduction in –log(*L*) for the age composition data, while the fit to the fishery CPUE actually got worse. Excluding the fishery CPUE data improved the fit to the fishery age and size composition data compared to both the model with longline CPUEs excluded and to the full model. The resulting estimates of biomass and ABC are considerably smaller and the biomass trajectory suggests a steeper decline between the mid-1980s and the most recent year, compared to models that include either the fishery or survey CPUE or both.

Excluding fishery size composition resulted in a similar overall fit in terms of the likelihood compared to the base model and a somewhat larger biomass and ABC. In general, the fit to the size composition data is relatively poor and the predicted size modes in 2000 and 2001 are shifted towards larger sizes relative to the observed modes. This apparent bias was evident in all model fits and it is not clear what causes it.

In addition to the base model, I ran a model that included one additional source of data, the survey size composition data. Although this is in some sense “double counting” the data, because length and age compositions are derived from the same data, I nevertheless fit a model with both size and age compositions included (1988–2006). The model fit was quite similar to the base model, suggesting that the size compositions are at least consistent with the other data. I did not include the survey size composition in any other models.

Sensitivity to Standard Error Estimates for MR Abundances

I examined whether the results are sensitive to the standard errors of the mark-recapture estimates by tripling the estimated standard errors (which are already twice the standard errors from the Petersen estimator). The result is similar to assuming a wider prior on *q_MR* and the resulting abundance estimates are larger than the mark-recapture estimates. The fit to most data components improved slightly. I mostly performed this check to see if the tight standard errors on the mark-recapture estimates force the declining trend over the last several years, which does not seem to be the case (while estimated abundances were generally larger for all years, the decreasing trend remained unchanged).

I ran several models that allowed “catchability” for the mark-recapture abundances to be freely estimated, rather than forcing q_{MR} to 1. The model provided a better fit to most data components, in particular the fishery age compositions and all abundance indices. The resulting biomass estimates were more than twice as large, but F_{40} was somewhat smaller (0.099) compared to the base model (0.105). As a result, the estimated ABC was 1,678 mt compared to 729 mt under the base model. This strongly suggests that a large proportion of the Chatham Strait sablefish population are not available to the tagging study or that the tagging study underestimates the total population size for other reasons.

Sensitivity to the Form of the Selectivity Relationship

Selectivity is fairly tightly constrained through the use of a simple 2-parameter logistic model. Freeing selectivity at present does not give sensible results (multiple domes in selectivity curve) and no alternative formulations were explored (e.g., dome-shaped selectivity, 3-parameter logistic).

Sensitivity to Assumed Variance of CPUE Estimates

For the ADMB implementation, I arbitrarily assumed that the CPUE estimates (longline survey and fishery) had coefficients of variation (CVs) of 20%. In the spreadsheet, I minimized squared differences on the log-scale, which is not a true likelihood and implies different weighting on these components than that implied by the likelihood used in the ADMB version. I did a simple sensitivity analysis to the assumed CV value by doubling the CV from 20% to 40% for either the longline survey or the fishery. This resulted in somewhat lower abundance estimates, but all of the likelihood components suggest a very similar fit to that achieved in the base model.

MODEL EVALUATION

Goodness of Fit and Diagnostics

Most models gave very reasonable results as evident in a comparison of observed and fitted values (Figure 55 to Figure 59) and as discussed above. The survey time series are reasonably well approximated although the fishery CPUE suggests stable abundances in recent years, whereas the base model (and most other models) suggests that abundances have continued to decline. This is obviously an important trend and should be closely monitored in coming years to evaluate whether abundances are indeed declining. The estimated trends can be compared to the estimated sablefish biomass and recruitment in the Gulf of Alaska/Bering Sea as estimated by NMFS (Figure 63, Figure 64). The trends are quite similar in that both show a pronounced peak in biomass in the mid- to late 1980s, but they also differ in important respects. In particular, the biomass trend in the Gulf of Alaska/Bering Sea does not show a continuing decline in biomass, unlike the apparent trend in Chatham Strait. Recruitment patterns also show similar long-term trends with strong recruitment in the late ‘70s/early ‘80s, a moderate increase in recruitment in the late 1990s and low recruitment in recent years (Figure 64). The remarkable similarity in the estimated trends suggest that the Chatham Strait population may indeed be part of the larger Gulf of Alaska/Bering Sea population and should be assessed as part of that population.

Age composition residuals show some obvious patterns (Figure 65, Figure 66). Most notably, the estimated fishery age composition does not account for the relatively large number of older fish observed in fishery samples, which may lead to an underestimate of total biomass. Although residuals are zero on average, if the number of older fish are underestimated and the number of younger age classes is overestimated (Figure 66), total biomass will be underestimated because of the larger weight of older fish.

Model Comparisons and Model Selection

I did not conduct a full and thorough comparison of alternative models, but results for eight alternative models are summarized in Table 6. Alternative models differ in terms of the weights (0 or 1) used for various likelihood components, which has the effect of including or excluding various components, and in terms of the width (*CV*) of the prior distributions on natural mortality (*M*) and catchability for the mark-recapture data (q_{MR}). The base case (Model 1) includes all data components and fixes $M = 0.1$ and $q_{MR} = 1$ (narrow priors). Models 2 and 3 relax the prior on *M* and result in estimates of *M* near the lower end of the probability distribution specified by the prior, suggesting that *M* is not well estimated. However, if both *M* and q_{MR} are estimated with a wide (non-constraining) prior (Models 4–8), reasonable estimates of both quantities are obtained with *M* ranging from 0.8 – 0.9 and a q_{MR} as low as 0.4 (implying that the mark-recapture estimate is only 40% of total biomass).

Excluding various data components when fitting the model (Models 5–8) substantially affected absolute biomass estimates, although the likelihood components showed only small differences. The overall time trend in biomass was very similar in all cases, but the absolute levels of biomass differed considerably among models. Based on the model fits, I saw little justification for excluding any data components from the model, even though the CPUE series did not seem to be particularly informative. I did not include a model that fixed $M = 0.1$ and allowed q_{MR} to be freely estimated, which would be a useful model to fit. However, the estimates of *M* are very reasonable and are similar to the most recent estimate of natural mortality for West Coast sablefish ($M = 0.08$).

A notable feature of all estimates is that the level of depletion (in terms of the ratio of the projected 2007 spawning biomass to the unfished spawning biomass) is very low. Depletion in models 1–8 ranged from 11–25% (Table 6), suggesting that biomass is well below the biomass that is expected to produce maximum sustainable yield in the long-term (B_{MSY}, which was not directly estimated but is often approximated by $B_{40\%}$). Table 6 also includes estimates of $F_{40\%}$ as well as the fishing mortality that would result from the NOAA harvest control rule under estimated levels of depletion (F_{ABC} (NOAA)).

CONCLUSIONS

The age-structured models considered here generally showed a very consistent biomass trend with a peak in sablefish biomass in Chatham Strait in the mid- to late 1980s, followed by a declining trend that either leveled off or reversed slightly around 1999/2000. Most of the model fits, especially when forced to fit the mark-recapture abundances, indicated a continuing decline in biomass in recent years, a trend that was not reflected in fishery CPUE (Figure 55), and was not paralleled in the Gulf of Alaska/Bering Sea stock (Figure 63), which otherwise showed a very similar trend. The general trend in biomass over time is very robust and the observed trend in fishery CPUE is reproduced even when the fishery CPUE data are excluded from the model fit. Therefore fishery CPUE appears to provide a reliable index of abundance and a much better relative index of abundance than the longline survey.

Different models differed greatly in the absolute estimate of biomass, which was substantially higher in models not tuned to mark-recapture data. Thus the absolute level of biomass is highly uncertain. Because there are no obvious sources of bias in the mark-recapture estimates and because these estimates were quite consistent among years, I recommend that the age structured assessment be constrained to fit the mark-recapture estimates until the issue of migration between Chatham Strait and outside waters is better understood or until a combined model can be fit to both the Chatham Strait and outside population of sablefish.

A comparison of the estimated biomass trends and recruitment patterns between Chatham Strait and the NOAA assessment of the Gulf of Alaska/Bering Sea stock of sablefish, combined with the evidence for extensive migrations between inside and outside waters, strongly suggests that they are part of a single population. Therefore, the best long-term option for obtaining unbiased estimates of management parameters for Chatham Strait sablefish is a combined stock assessment that incorporates migration between these areas.

RECOMMENDATIONS FOR AGE-STRUCTURED ANALYSIS

1) As indicated under overall recommendations, I recommend that ADF&G develop a full age-structured model for sablefish in Chatham Strait to integrate various sources of existing data and to reconstruct historical trends, evaluate current stock status, and estimate management parameters. In particular, $F_{40\%}$ and the associated Allowable Biological Catch can be estimated directly within the model to set quotas.

2) Because the estimates of absolute abundance and biomass are highly uncertain, I recommend that the catchability for the mark-recapture estimates of abundance (q_{MR}) be fixed at 1, at least until movement patterns of sablefish between Chatham Strait and outside waters and associated biases are better understood.

3) Once the base model is developed and has been sufficiently tested (e.g., the base model as implemented in ADMB), I recommend that the stock assessment analyst explore a small set of reasonable models each year (3–4) within a general modeling framework. As resources and time allow, incremental improvements to the model should be pursued as new issues arise.

4) The importance of tagging, longline, and fishery sampling data to ASA should ultimately be explored using management strategy evaluation (see, for example http://www.cmar.csiro.au/research/mse/). Surveys could then be improved, survey frequency could be reduced, or surveys may even be discontinued as appropriate. Limited sensitivity analyses suggest that no single data source has a clearly dominant effect on model results.

5) After some experience with the age-structured model has been gained and the model has become accepted, gradual improvements to the model should be explored as time and

resources allow. These may include, but are not limited to, some of the following modifications:

a) Both the fishery and CPUE indices could be refined, for example through the use of a General Linear Model to estimate annual average CPUE anomalies while accounting for differences in sampling design and gear configurations among years.

b) Because of the differences in growth between male and female sablefish, the model could be expanded to a split-sex model (where growth is modeled separately for males and females for computing total biomass)

c) The inclusion of a stock-recruitment relationship in the model could be explored to improve recruitment estimates and predictions of future biomass.

d) The current model includes an aging error matrix that is identical to the one used in the NOAA sablefish model and is based on aging done at the Alaska Fisheries Science Center (AFSC). Because of differences in aging between the ADF&G and AFSC aging labs, the age-transition matrix should be verified against results obtained by the ADF&G aging lab.

e) The model currently assumes asymptotic selectivity in both the longline survey and the fishery. Alternative selectivity curves (more flexible shape, including dome-shaped) could be explored to obtain a better fit to the age data.

f) Growth and maturity schedules should be evaluated periodically to monitor for changes over time that would bias estimates.

g) Sensitivity analyses to examine the impact of potential changes in weight-at-age over time on model results should be conducted.

6) While the age-structured analysis would, in my view, provide a useful and improved approach to determining acceptable catch, there are clearly important, unresolved issues that cannot be adequately addressed using an ASA for Chatham Strait alone. In particular, there is abundant evidence for extensive migrations between Chatham Strait and outside waters and it is possible, and even likely, that Chatham Strait sablefish and sablefish in the Gulf of Alaska and Bering Sea/Aleutians are part of a single stock. Therefore I strongly recommend that ADF&G work with NOAA and other agencies to better understand the magnitude and importance of migration between Chatham Strait and outside waters. It is my impression that there is a wealth of tagging data collected by both State and federal agencies, these data should be compiled and analyzed to better resolve migration patterns.

7) Based on current knowledge alone, it is clear that the Chatham Strait sablefish population is not an isolated population. Therefore, in addition to improving our understanding of sablefish migration, I recommend that ADF&G work with NOAA and DFO to pursue the development of a single stock assessment for all sablefish in the Gulf of Alaska and adjacent waters that accounts for migration between different areas. It is my understanding that NOAA is currently developing a spatially explicit age-structured model that includes migration. The development of an age-structured model for Chatham Strait would be a helpful step towards

modeling Chatham Strait sablefish as a subpopulation of the larger Gulf of Alaska/Bering Sea population that is linked with this population via migration. However, a combined model for State and federal waters would require an agreement between federal and State agencies as to the allocation of allowable catches between federal and State waters.

ACKNOWLEDGEMENTS

I would like to acknowledge Sherri Dressel, Alaska Department of Fish and Game Groundfish and Herring Biometrician and Biometrics Supervisor for her contribution of time and expertise to provide background, data, and facilitate the analysis. I would like to thank Kyle Hebert, Marine Fisheries Supervisor, for supporting and administering the contract with Alaska Department of Fish and Game under which this work was completed. I also acknowledge Dave Carlile who previously held the Groundish Biometrician position and who constructed the original ASA model used for Chatham Strait sablefish.

REFERENCES CITED

Carlile, D. W., B. Richardson, M. Cartwright, and V. M. O'Connell. 2002. Southeast Alaska sablefish stock assessment activitites 1988-2001. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 1J02-02, Juneau.

Clark, W.G. 1991. Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries and Aquatic Sciences 48: 734–750.

Clark, W.G. 1993. The effect of recruitment variability on the choice of target level of spawning biomass per recruit. [*in*] Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. Edited by G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn, II. Alaska Sea Grant College Program. pp. 233–246.

Clark, W.G., Hare, S.R., Parma, A.M., Sullivan, P.J., and Trumble, R.J. 1999. Decadal changes in growth and recruitment of Pacific halibut (*Hippoglossus stenolepis*). Canadian Journal of Fisheries and Aquatic Sciences 56: 242–252.

Clausen, D.M., Fujioka, J.T., and Heifetz, J. 1997. Estimates of sablefish, *Anoplopoma fimbria,* abundance in the Gulf of Alaska based on depletion experiments using longlines. [*in*] Biology and management of sablefish. Edited by M. Wilkins and M. Saunders. NOAA Tech. Rep. NMFS 130. pp. 171–182.

Dichmont, C., Punt, A., Deng, A., Dell, Q., and Venables, W. 2003. Application of a weekly delay-difference model to commercial catch and effort data for tiger prawns in Australia's northern prawn fishery. Fisheries Research 65: 335–350.

Haist, V., Hilborn, R., and Wyeth, M. 2001. Sablefish stock assessment for 2001 and advice to managers for 2002. Department of Fisheries and Oceans. Canadian Science Advisory Research Document No. 2001/135.

Haist, V., Kronlund, A.R., and Wyeth, M.R. 2005. Sablefish (*Anoplopoma fimbria*) in British Columbia, Canada: stock assessment for 2004 and advice to managers for 2005. Department of Fisheries and Oceans. Canadian Science Advisory Research Document No. 2005/031.

Hanselman, D.H., Lunsford, C.R., Fujioka, J.T., and Rodgveller, C.J. 2006. Alaska sablefish assessment for 2007. [*in*] Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, 605 W. 4th Ave., Suite 306, Anchorage, AK 99501.

REFERENCES CITED (Continued)

Hanselman, D.H., Lunsford, C.R., Fujioka, J.T., and Rodgveller, C.J. 2007. Alaska sablefish assessment for 2008. [*in*] Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, 605 W. 4th Ave., Suite 306, Anchorage, AK 99501. pp. 195–312.

Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman and Hall, New York.

Maloney, N.E., and Heifetz, J. 1997. Movements of tagged sablefish, *Anoplopoma fimbria*, released in the eastern Gulf of Alaska. [*in*] Biology and management of sablefish. Edited by M. Wilkins and M. Saunders. NOAA Tech. Rep. NMFS 130. pp. 115–121.

National Research Council. 1998. Improving fish stock assessments. National Academy Press, Washington, D.C.

NPFMC. 2007. Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions, North Pacific Fishery Management Council, 605 W. 4th Ave., Suite 306, Anchorage, AK 99501.

Otter Research. 2007. An introduction to AD Model Builder Version 8.0.2. Otter Research Ltd., Box 2040, Sidney, British Columbia, Canada.http://otter-rsch.ca/adm.zip

Parma, A.M. 1990. Optimal harvesting of fish populations with non-stationary stock-recruitment relationships. Natural Resource Modeling 4(1): 39–76.

Parma, A.M., and Deriso, R.B. 1990. Experimental harvesting of cyclic stocks in the face of alternative recruitment hypotheses. Canadian Journal of Fisheries and Aquatic Sciences47 (3): 595–610.

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

Schirripa, M.J., and Colbert, J.J. 2005. Status of the sablefish resource off the continental U.S. Pacific coasts in 2005. Pacific Fishery Management Council, Portland, Oregon.

Schirripa, M.J., and Colbert, J.J. 2006. Interannual changes in sablefish (*Anoplopoma fimbria*) recruitment in relation to oceanographic conditions within the California Current System. Fisheries Oceanography 15(1): 25–36.

Seber, G.A.F. 1982. The estimation of animal abundance. MacMillan Publishing Co., New York.

Sigler, M.F. 1999. Estimation of sablefish, *Anoplopoma fimbria*, abundance off Alaska with an age-structured population model. Fishery Bulletin 97: 591–603.

Smith, G.B. 1981. The biology of walleye pollock. In The Eastern Bering Sea Shelf: Oceanography and Resources. Edited by D.W. Hood and J.A. Calder. U.S. Department of Commerce.

Spiegelhalter, D.J., Thomas, A., Best, N.G., and Gilks, W.R. 1995. BUGS: Bayesian inference Using Gibbs Sampling, Version 0.50. MRC Biostatistics Unit, Cambridge.

Stahl, J. 2004. Maturation of walleye pollock, *Theragra chalcogramma*, in the Eastern Bering Sea in relation to temporal and spatial factors. M.S., School of Fisheries and Ocean Sciences, University of Alaska Fairbanks.

Thompson, G.G., Dorn, M.W., Gaichas, S., and Aydin, K. 2006. Chapter 2: Assessment of the Pacific Cod Stock in the Eastern Bering Sea and Aleutian Islands Area. [*in*] Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, 605 W. 4th Ave., Suite 306, Anchorage, AK 99501.

Walters, C., and Parma, A.M. 1996. Fixed exploitation rate strategies for coping with effects of climate change. Canadian Journal of Fisheries and Aquatic Sciences 53: 148–158.

Walters, C.J., and Martell, S.J.D. 2004. Fisheries Ecology and Management. Princeton University Press, Princeton.

Williams, B.K., Nichols, J.D., and Conroy, M.J. 2002. Analysis and Management of Animal Populations. Academic Press, San Diego, U.S.A.

TABLES

Table 1.–Statistical areas where tagged sablefish were released and recovered, and percentage of sablefish tagged in a given release area that were recovered in a different area.

Recovery	**Release area**							
area	**335701**	**345603**	**345631**	**345701**	**345702**	**345731**	**345803**	**Total**
335701	1	0	3	1	0	2	0	7
345603	0	10	9	3	0	2	1	25
345631	0	11	107	48	2	2	3	173
345701	0	5	20	139	1	11	3	179
345702	0	0	4	3	1	0	0	8
345731	0	1	5	7	0	12	5	30
345803	0	0	5	4	0	2	7	18
355801	0	0	0	1	0	0	0	1
Unknown	1	4	12	28	1	11	3	60
Total	2	31	165	234	5	42	22	501
Total known	1	27	153	206	4	31	19	441
% moved	0.0	63.0	30.1	32.5	75.0	61.3	63.2	37.2

Table 2.–Notation for models used in mark-recapture estimation.

Notation	Definition
N_0	Number of sablefish in Chatham Strait at the time of marking
M_0	Number of marks released
D	Known number of marks removed that are not available to either the LL survey or to the fishery (tags from halibut fishery, from outside, etc.)
B	Number of fish entering Chatham Strait between pot survey and LL survey (unknown)
m	Annual instantaneous rate of natural mortality (may include emigration)
μ	Daily instantaneous rate of natural mortality (may include emigration) = $m/365$
δm	Difference in mortality rate between marked and unmarked fish, hence instantaneous mortality of marked fish is $(m + \delta m)$.
i	Subscript for time period i, which may refer to longline survey (i = 1) or to one of 18 consecutive 5-day fishery periods (based on time of landing, $i = 2,\ldots,18$)
N_i	Number of sablefish in Chatham Strait at the beginning of time period i
M_i	Number of marked sablefish in Chatham Strait at the beginning of time period i
t_i	Total number of days between beginning of period i-1 and beginning of period i
C_i	Total catch (number of sablefish removed) reported during period i
n_i	Observed catch (number of marked + unmarked sablefish that were checked for clips) during period i ($n_i < C_i$)
m_i	Number of marked fish recovered during period i
β	Number of unmarked fish entering Chatham Strait from outside per day
p_i	Proportion of marked fish in the population at beginning of period i (= M_i/n_i)
q	Catchability coefficient for the fishery relating fishery CPUE in period i to the abundance of sablefish $CPUE_i = q * N_i$

Table 3.–Models fit to 2006 sablefish data for mark-recapture estimation of abundance.

Model	Description	Parameters
A	Petersen estimator	N
B	Time-stratified estimator with natural mortality	N, m
C	Time-stratified, natural mortality, and immigration	N, β, m
D	Time-stratified estimator with natural mortality including fishery CPUE data	N, m, q, σ
E	Time-stratified estimator with natural mortality and immigration, including fishery CPUE data	N, β, m, q, σ

Notes:

1. Natural mortality was fixed at 0.1 or highly constrained in all models
2. Average abundance across the fishing season ($\overline{N}$) was estimated in models A, C, and E, whereas initial abundance at the beginning of the season (N_1) is estimated by models B and D.

Table 4.–Summary of model results and model comparisons.

Model	A	B	C	D	E
Immigration	No	No	Yes	No	Yes
fishery CPUE	No	No	No	Yes	Yes
natural mortality	No	Yes	Yes	Yes	Yes
Parameter estimates [a]					
N (mean numbers of fish)	2,428,000	2,492,000	2,477,000	2,487,000	2,364,000
m	0	0.1	0.1	0.1	0.1
Immigration (β) (numbers of fish)			11,950		5,858
catch coeff (q)				4.01E-05	3.97E-05
se(CPUE) (σ)				55.88	61.66
Posterior distribution of abundance [b]					
2.5th percentile	2,214,000	2,289,000	2,227,000	2,288,000	2,145,000
5th percentile	2,249,000	2,320,000	2,262,000	2,319,000	2,176,000
Mean	2,428,000	2,492,000	2,477,000	2,487,000	2,364,000
95th percentile	2,607,000	2,675,000	2,713,000	2,668,000	2,570,000
97.5th percentile	2,641,000	2,711,000	2,763,000	2,706,000	2,618,000
std error	108,933	107,700	137,200	107,100	119,800
Maximum likelihood components (negative log-likelihoods)					
MR data		58.984	54.627	58.984	56.022
fishery CPUE				74.807	76.003
Total likelihood		58.984	54.627	133.790	132.025
# of parameters		1	2	3	4
sample size		19	19	37	37
AIC [c]		120.2026	114.0036	274.3079	273.3008
LRT [d]			8.713751		3.529766
p-value [d]			0.003158		0.060276

Notes:

[a] Model A: Petersen estimator. Models B–E: Bayesian model estimates with informative prior on *m*.

[b] Percentiles for Model A based on normal distribution and standard error.

[c] AIC: Small-sample Akaike Information Criterion for model comparisons

[d] Likelihood-ratio test statistic for comparing models B vs. C and D vs. E, respectively, with associated *p*-values

Table 5.–Number of clipped fish and selectivity at size estimated from 2006 NOAA sablefish assessment (Hanselman et al. 2006). Size classes denote the center of each 20mm size class.

Size class	Numbers of clipped fish	NOAA fishery selectivity
475	0	0.015
495	1	0.027
515	223	0.049
535	352	0.096
555	352	0.191
575	438	0.369
595	541	0.605
615	658	0.800
635	739	0.914
655	664	0.969
675	605	0.991
695	478	0.998
715	338	1.000
735	239	1.000
755	131	1.000
775	101	1.000
795	66	1.000
815	48	1.000
835	29	1.000
855	7	1.000
875	17	1.000
895	17	1.000
915	10	1.000
935	5	1.000
955	5	1.000
975	3	1.000
995	3	1.000
1015	0	1.000
1035	2	1.000

Table 6.–Input weights, selected parameter estimates, other quantities of interest, and likelihood components for eight alternative models, which differ in weights and in assumed priors for *M* and q_{MR}. Models 1–4 use all data components, models 4–8 use wide priors on both *M* and *q*.

	Fixed *M*	**Free *M***		**Free *q* & *M***	**Excluding components**			
	Base case	**Narrow prior**	**Wide prior**	**Wide priors**	**Longline CPUE**	**Fishery CPUE**	**Longline and fishery**	**NO fishery data**
Model	1	2	3	4	5	6	7	8
Weights:								
Recruitment dev	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Fishing mort dev	0	0	0	0	0	0	0	0
Survey CPUE 1	1	1	1	1	0	1	0	1
Survey CPUE 2	1	1	1	1	0	1	0	1
Fishery CPUE	1	1	1	1	1	0	0	0
Mark-recapture abundance	1	1	1	1	1	1	1	1
Fishery age comp	1	1	1	1	1	1	1	0
Survey age comp	1	1	1	1	1	1	1	1
Fishery size comp	1	1	1	1	1	1	1	0
Survey size comp	0	0	0	0	0	0	0	0
Parameters & priors								
M (fixed or estimated)	0.100	0.078	0.061	0.077	0.080	0.081	0.089	0.084
M prior	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
M prior CV	0.01	0.1	0.2	0.2	0.2	0.2	0.2	0.2
q_MR (fixed or estimated)	1.0	1.0	1.0	0.527	0.581	0.389	0.750	0.437
q_MR prior	1	1	1	1	1	1	1	1
q_MR prior CV	0.01	0.01	0.01	0.4	0.4	0.4	0.4	0.4

-continued-

Table 6.–Page 2 of 2.

	Fixed *M*	**Free *M***		**Free *q* & *M***	**Excluding components**			
	Base case	**Narrow prior**	**Wide prior**	**Wide priors**	**Longline CPUE**	**Fishery CPUE**	**Longline and fishery**	**NO fishery data**
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Derived quantities								
$F_{35\%}$	0.128	0.095	0.073	0.091	0.094	0.104	0.102	0.096
$F_{40\%}$	0.105	0.078	0.061	0.075	0.078	0.086	0.084	0.080
$F_{50\%}$	0.071	0.054	0.042	0.052	0.054	0.059	0.058	0.055
F_{ABC} (NMFS)	0.033	0.023	0.014	0.036	0.033	0.051	0.033	0.044
SSB_0	26.911	26.925	32.359	34.460	33.206	39.095	26.103	38.433
$SSB_{40\%}$	10.765	10.770	12.944	13.784	13.282	15.638	10.441	15.373
ABC	0.725	0.546	0.429	1.026	0.932	1.465	0.720	1.392
2006 age 4+ biomass	8.042	7.722	7.625	14.170	12.587	19.437	9.260	17.652
2006 exploitable biomass	7.341	7.321	7.341	13.826	12.351	19.039	9.169	16.965
2006 spawn biomass	4.067	3.868	3.796	7.118	6.369	9.969	4.742	9.188
2007 age 4+ biomass	7.077	6.877	6.861	13.231	11.591	18.167	8.336	16.548
2007 exploitable biomass	6.629	6.624	6.685	12.998	11.477	17.918	8.331	16.035
2007 spawn biomass	3.727	3.577	3.544	6.883	6.078	9.688	4.383	8.891
Depletion	0.138	0.133	0.110	0.200	0.183	0.248	0.168	0.231
Likelihoods								
Survey 1 CPUE	12.411	11.741	11.183	10.560		9.576		9.042
Survey 2 CPUE	6.833	6.042	5.283	2.568		1.776		1.681
Fishery CPUE	11.218	9.213	7.630	6.122	6.138			
MR abundance	4.927	4.735	4.541	4.139	3.900	4.265	3.655	4.313
Fishery age comp	20.716	17.536	16.054	13.635	13.598	16.616	13.511	
Survey age comp	37.847	37.015	37.186	38.073	37.785	36.686	38.116	35.279
Fishery size comp	8.000	8.198	7.778	8.201	8.279	5.683	7.543	
TOTAL Data	101.952	94.479	89.655	83.297	69.699	74.603	62.825	50.314
Other parameters								
*q*_Survey 1 CPUE	0.020	0.022	0.025	0.018		0.016		0.019
*q*_Survey 2 CPUE	0.091	0.095	0.099	0.059		0.045		0.051
*q*_fishery	0.025	0.027	0.029	0.020	0.021			0.025
avgR (1980–2006)	0.912	0.614	0.496	0.776	0.792	0.926	0.740996	0.981

FIGURES

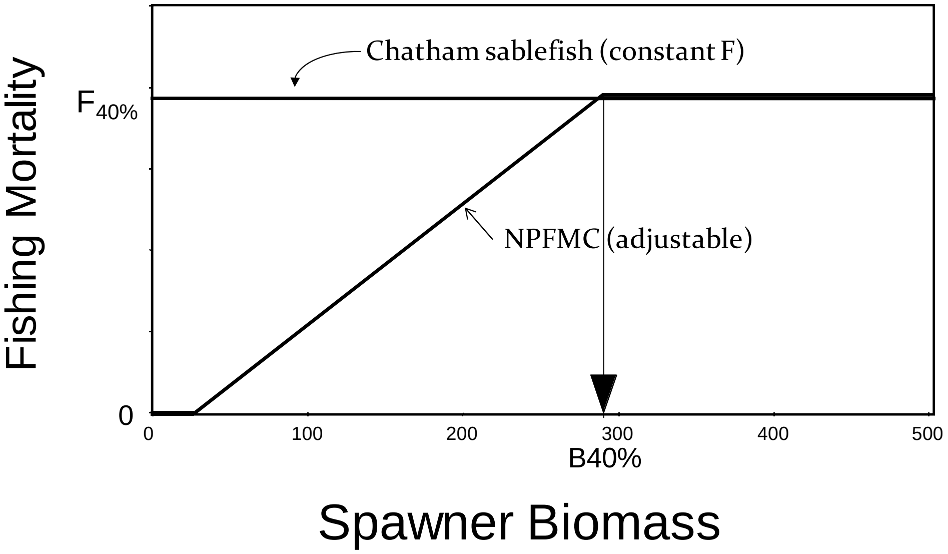


Figure 1.–Two alternative harvest policies and control rules: fixed harvest rate policy as applied to Chatham sablefish and adjustable rate policy as applied by NPFMC to most groundfish stocks in federal waters off Alaska.

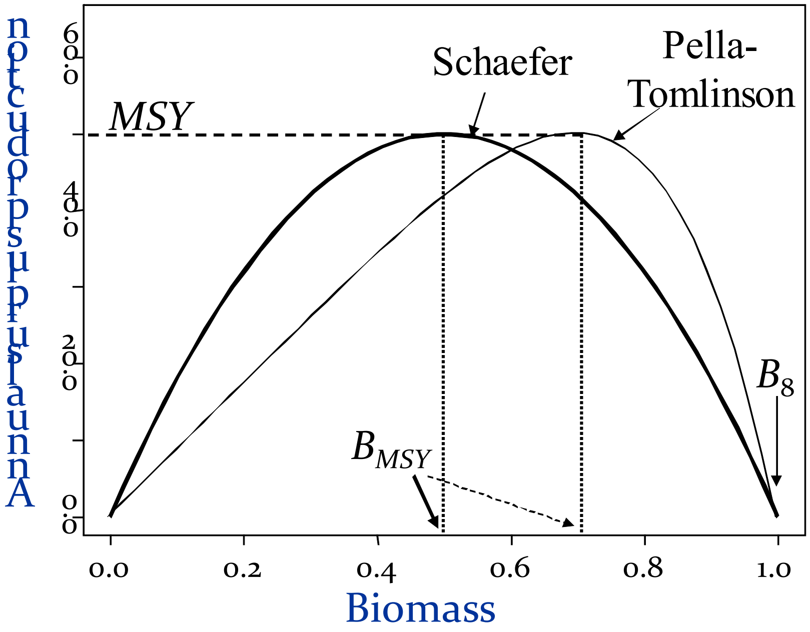


Figure 2.–Examples of surplus production models, including Schaefer model and one example of the Pella-Tomlinson model.

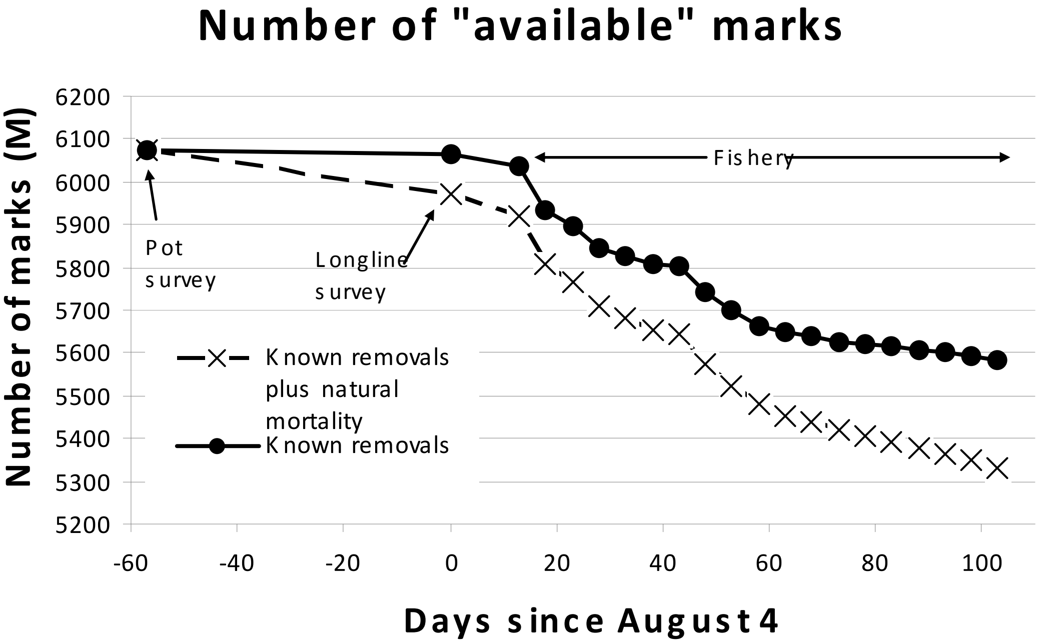


Figure 3.–Number of marked fish remaining in the population after accounting for known removals only and after accounting for known removals plus natural mortality (assuming M=0.1).

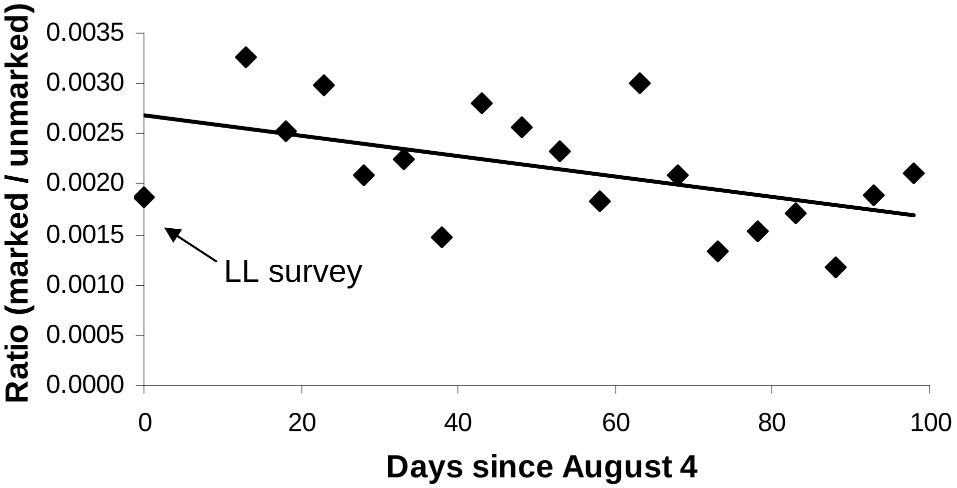


Figure 4.–Change in the ratio of marked to unmarked fish caught in the fishery over the course of the fishing season with linear fit (F1,17 = 5.542, p = 0.0309, R2 = 0.26). Ratios were computed for the longline survey and for each of 18 consecutive 5-day periods during the fishery, based on delivery date.

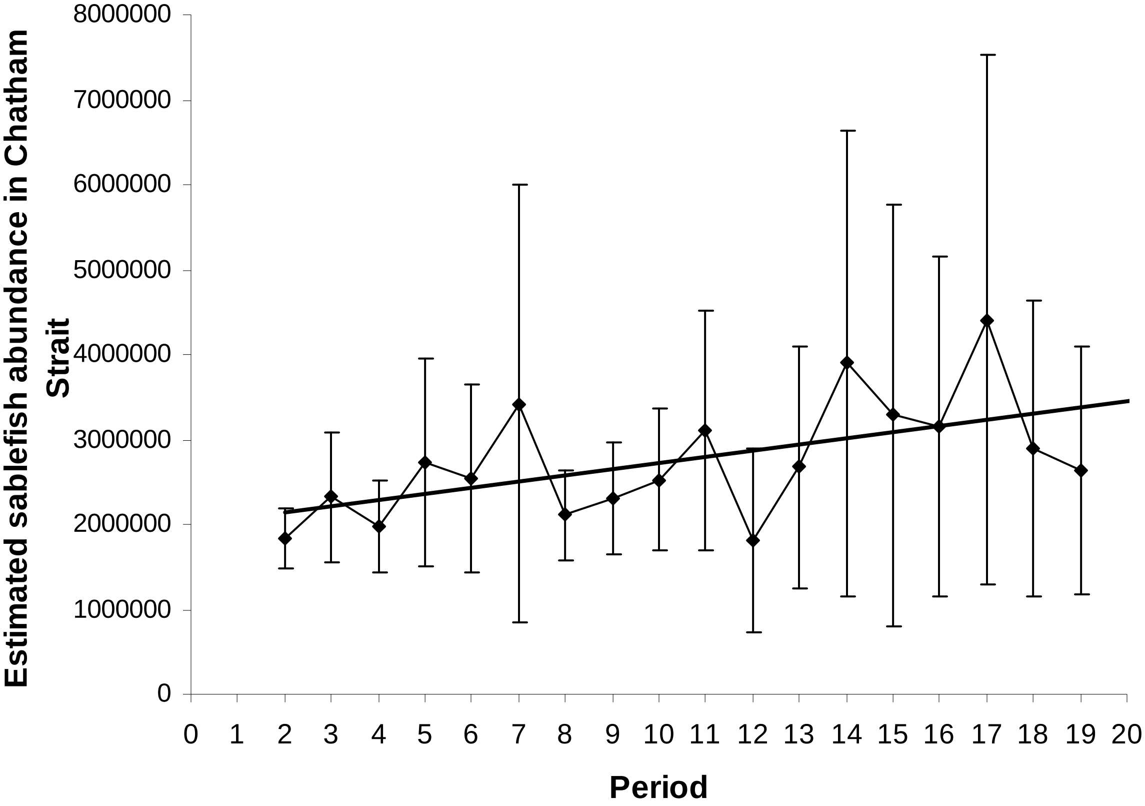


Figure 5.–Time-stratified Petersen estimates of sablefish abundance in Chatham Strait using marked and unmarked sablefish counted during consecutive 5-day periods in the fishery with approximate 95% confidence intervals and linear time trend (weighted least-squares regression).Estimates and confidence intervals for each period are based on Chapman modification of the Petersen estimator and its variance.

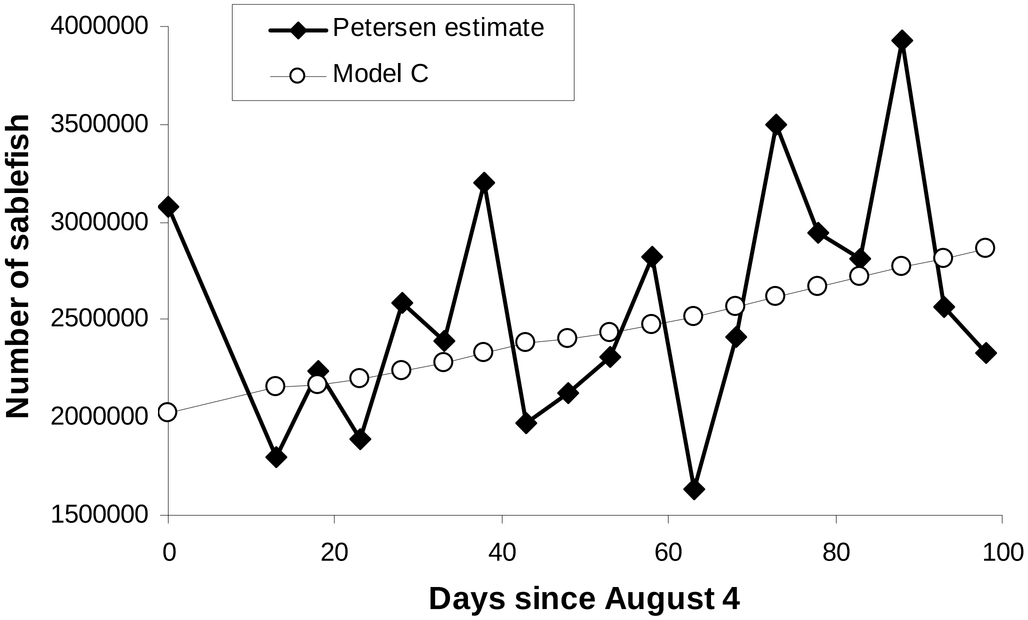


Figure 6.–Time-stratified Petersen estimates of sablefish abundance in Chatham Strait with model estimates, which include a daily constant immigration of sablefish into Chatham Strait.

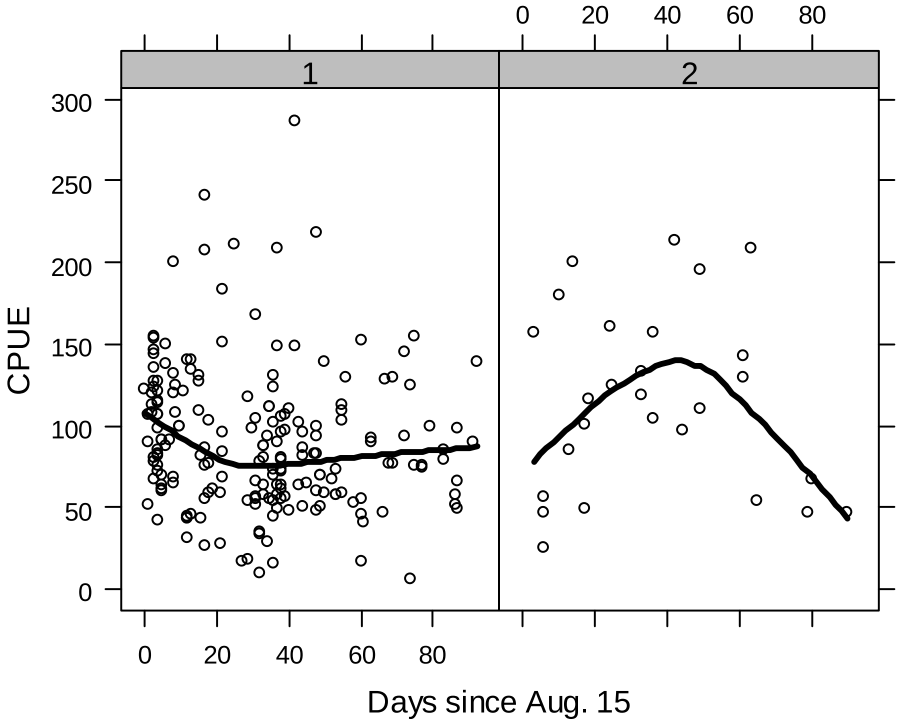


Figure 7.–CPUE of sablefish (number of sablefish per 1,000 hooks) for each reported fishing trip in the Chatham Strait fishery by landing date and gear type (1 indicates conventional gear, 2 indicates snap-on gear) with LOWESS smooth.

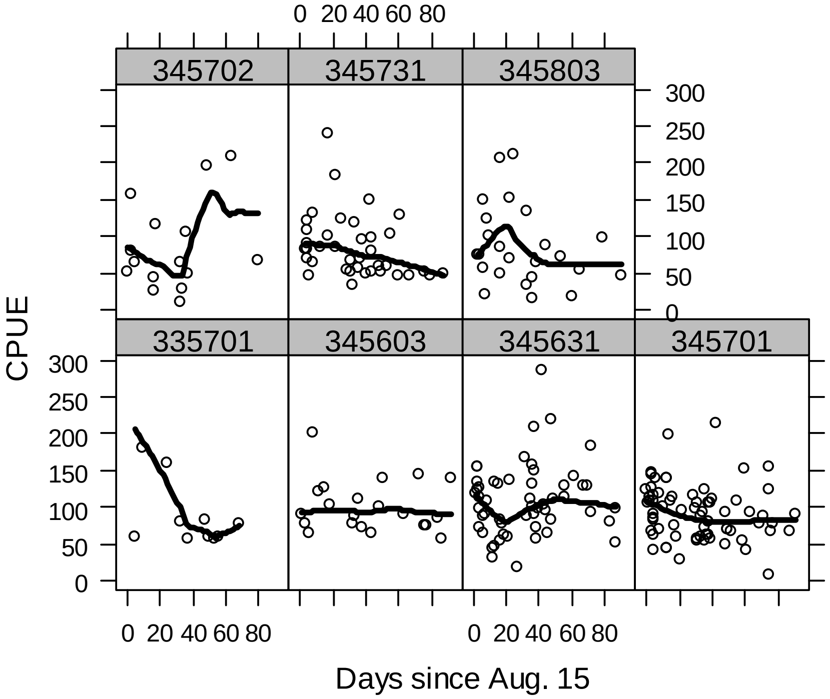


Figure 8.–CPUE of sablefish (number of sablefish per 1,000 hooks) for each reported fishing trip in the Chatham Strait fishery by landing date and statistical area with LOWESS smooth.

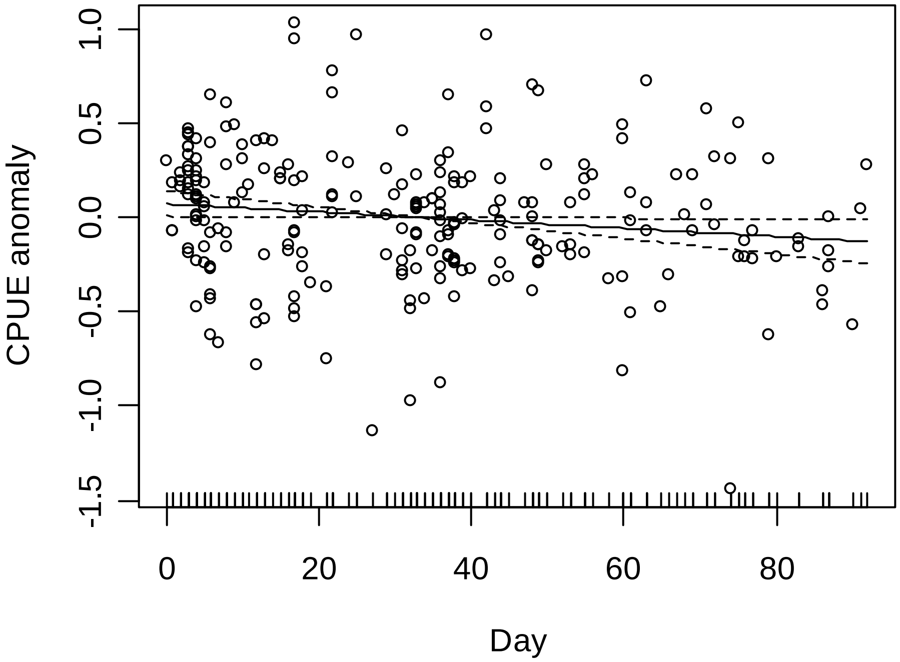


Figure 9.–CPUE of sablefish (pounds/hook, fourth-root transformed) for each reported fishing trip in the Chatham Strait fishery by landing date (circles) with best non-parametric regression fit (GAM, best model was a linear model) (solid line) and approximate 95% confidence band (dashed lines).

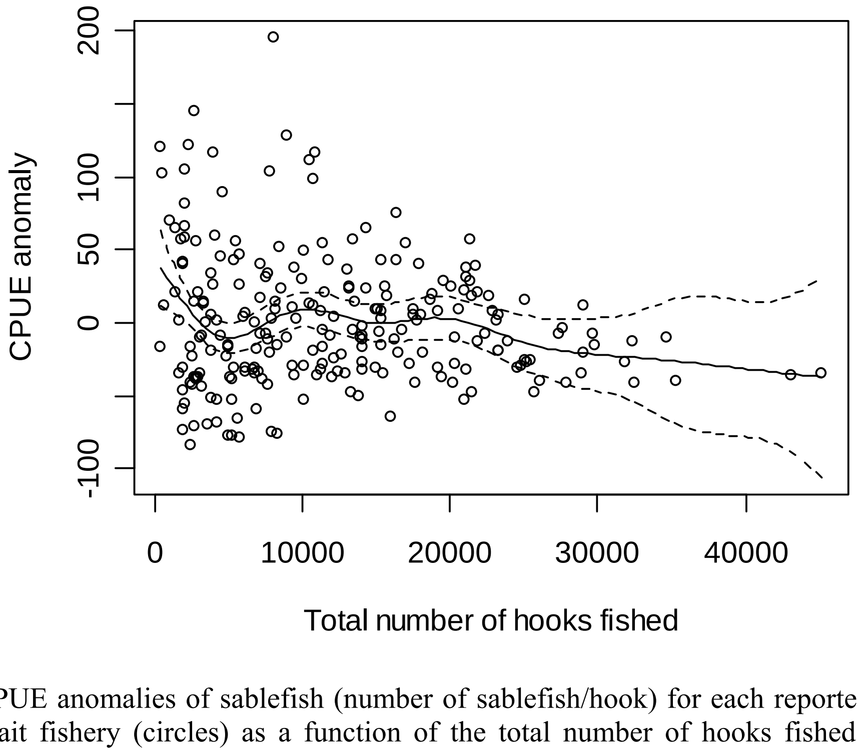


Figure 10.–CPUE anomalies of sablefish (number of sablefish/hook) for each reported fishing trip in the Chatham Strait fishery (circles) as a function of the total number of hooks fished with best non-parametric regression fit (GAM) (solid line) and approximate 95% confidence band (dashed lines). The estimate accounts for decrease in CPUE over time [GAM model: CPUE ~ s(Day) + s(# of hooks)].

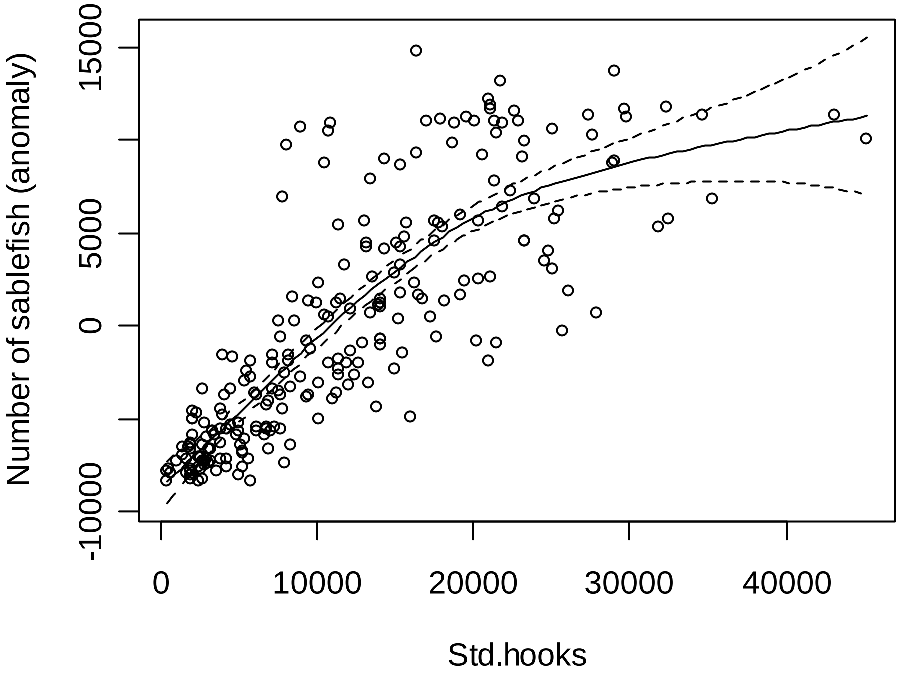


Figure 11.–Number of sablefish caught per trip as a function of number of standardized hooks (evidence of saturation above approximately 20,000 hooks).

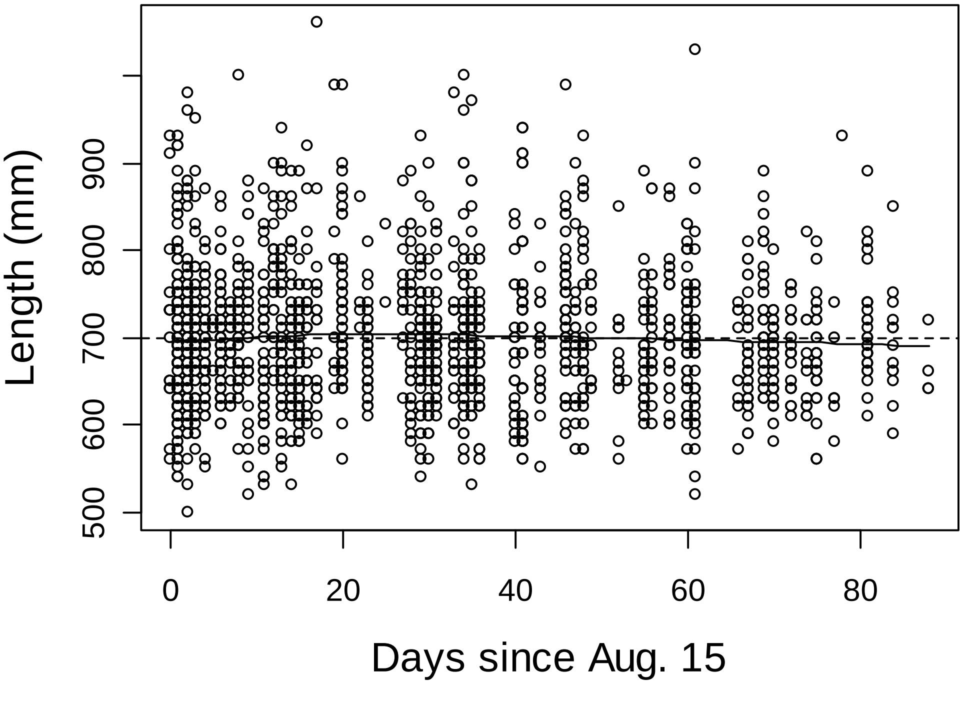


Figure 12.–Scatter plot of lengths of sablefish by delivery day with LOWESS smooth (solid line) and overall mean length (dashed line). No significant trend in length over time (GAM: $p = 0.330$).

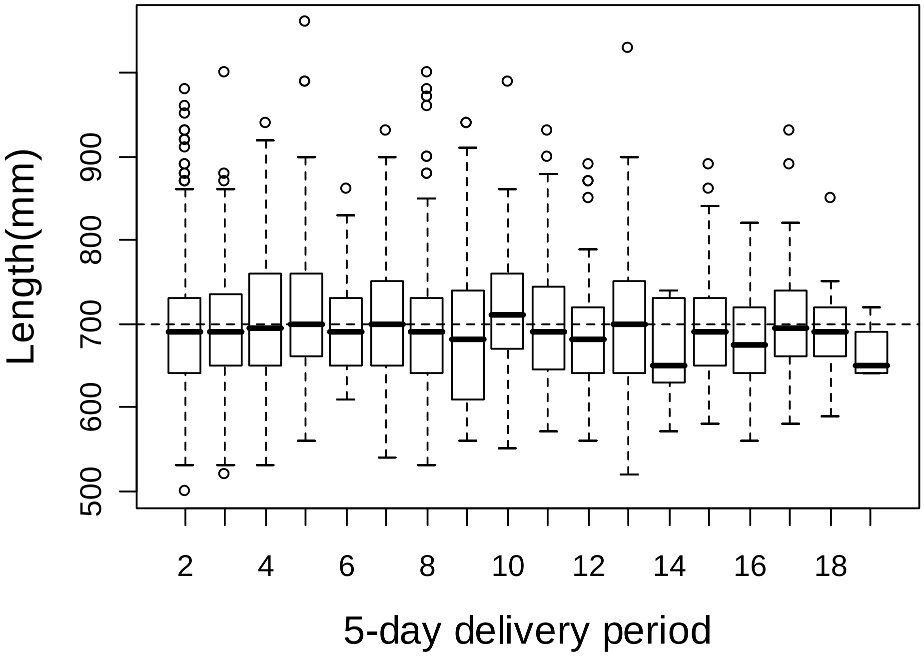


Figure 13.–Box plots of lengths of sablefish by delivery period with overall median length (dashed line). No significant difference in lengths among periods (one-way ANOVA: p = 0.173).

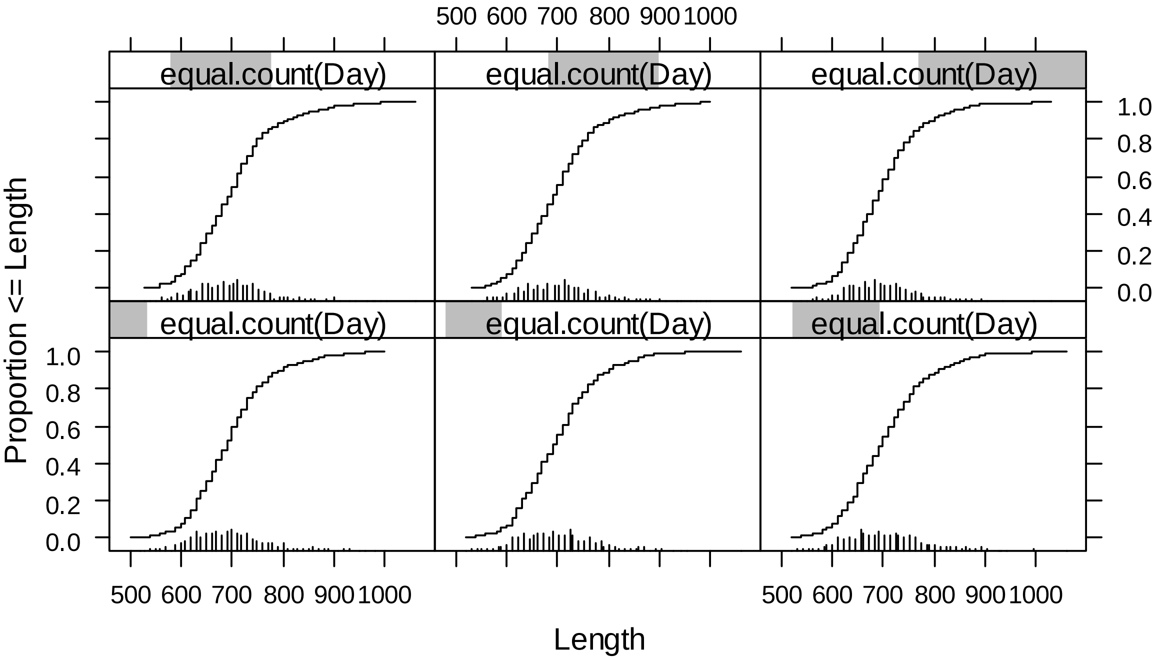


Figure 14.–Cumulative length-frequencies and stick histograms of sablefish sampled from the fishery by delivery day. Delivery days were split into 6 groups with approximately equal number of sampled fish per group and overlap among groups (as indicated by grey bars).

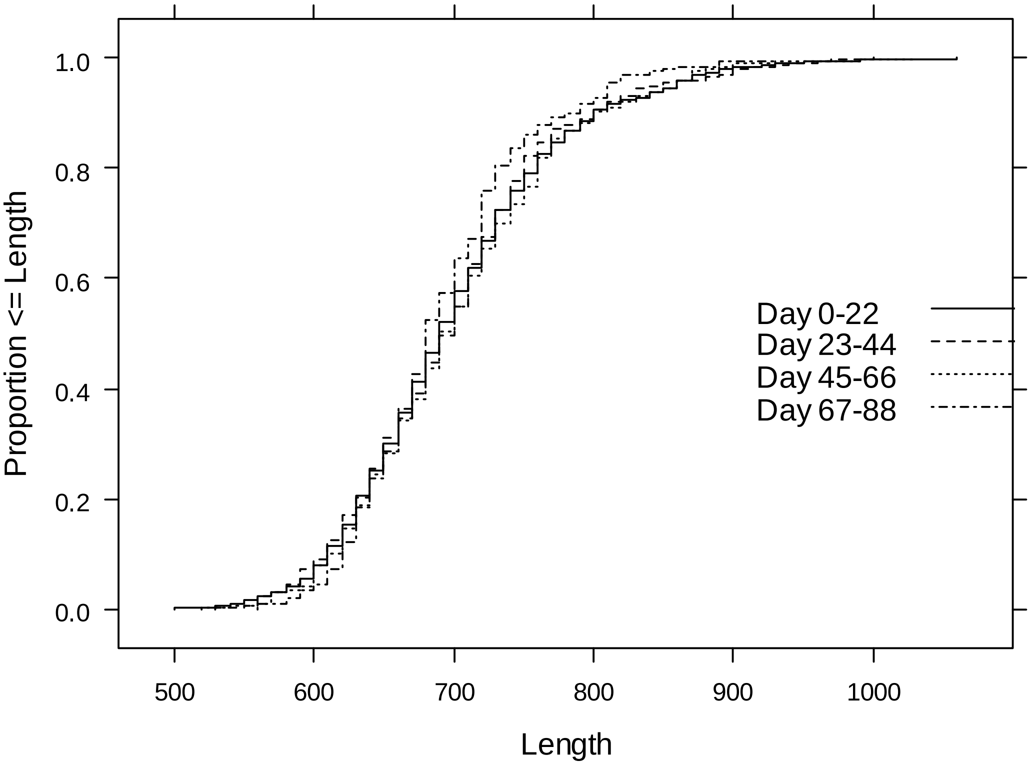


Figure 15.–Cumulative length-frequencies of sablefish sampled from the fishery for four 22-day periods (based on delivery day).

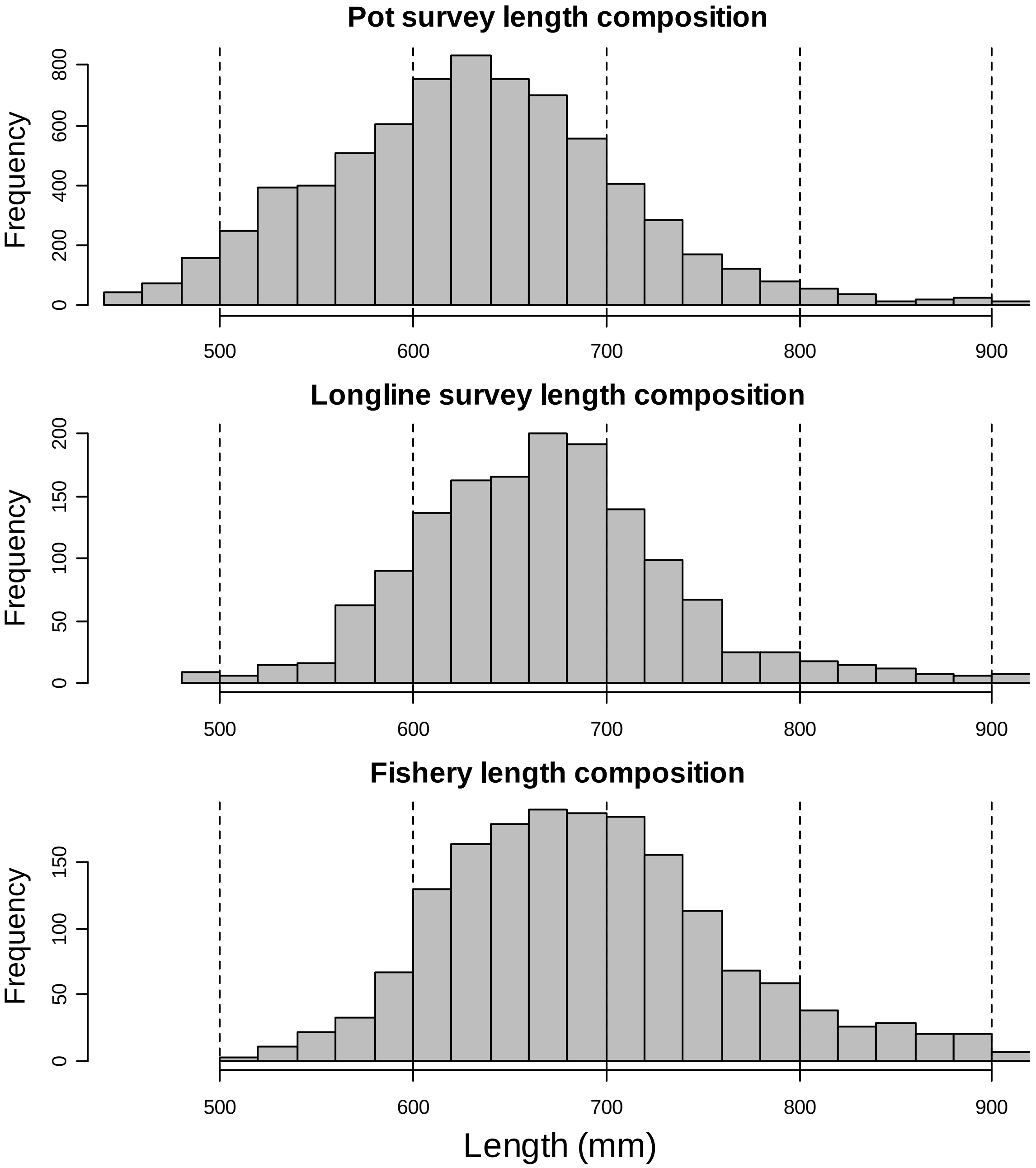


Figure 16.–Histograms of the lengths of sablefish caught in pot survey, longline survey, and fishery in 2006.

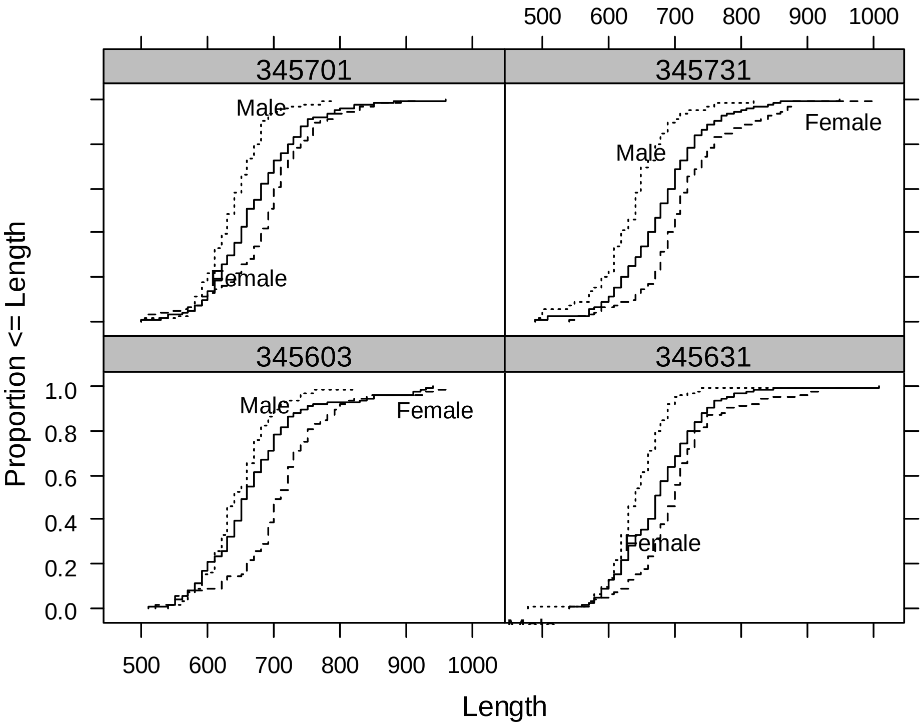


Figure 17.–Cumulative length-frequencies of unsexed (solid line) and of male (dotted) and female (dashed) sablefish sampled during longline survey by ADF&G statistical area.

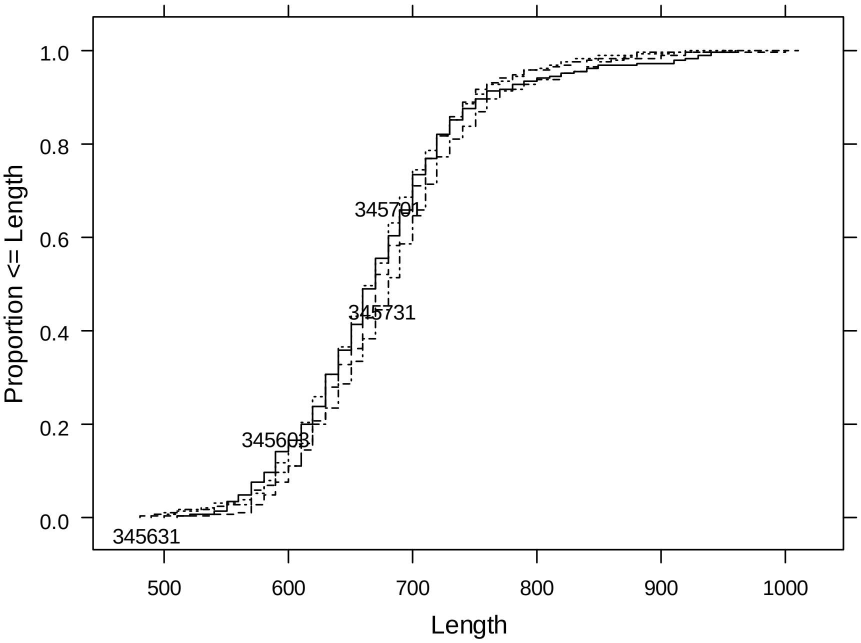


Figure 18.–Cumulative length-frequencies of all sablefish sampled during longline survey by ADF&G statistical area.

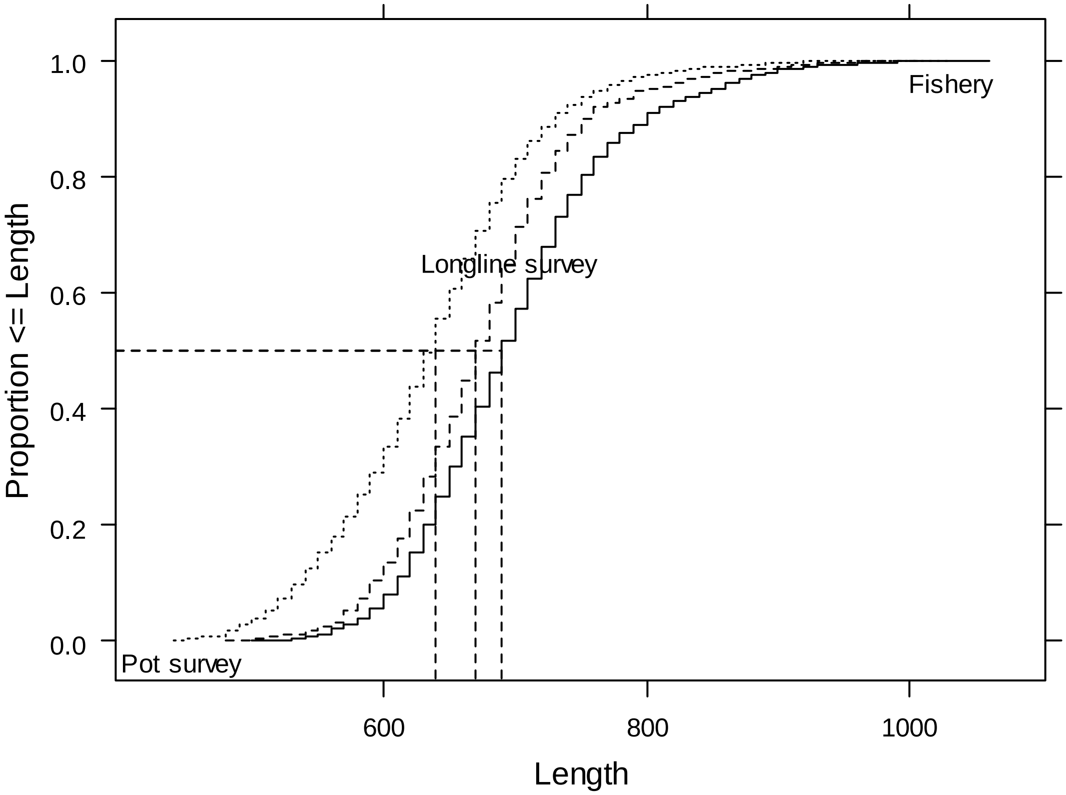


Figure 19.–Cumulative length-frequencies of all sablefish sampled during pot survey (dotted), longline survey (dashed), and fishery (solid) in 2006 (50th percentile indicated).

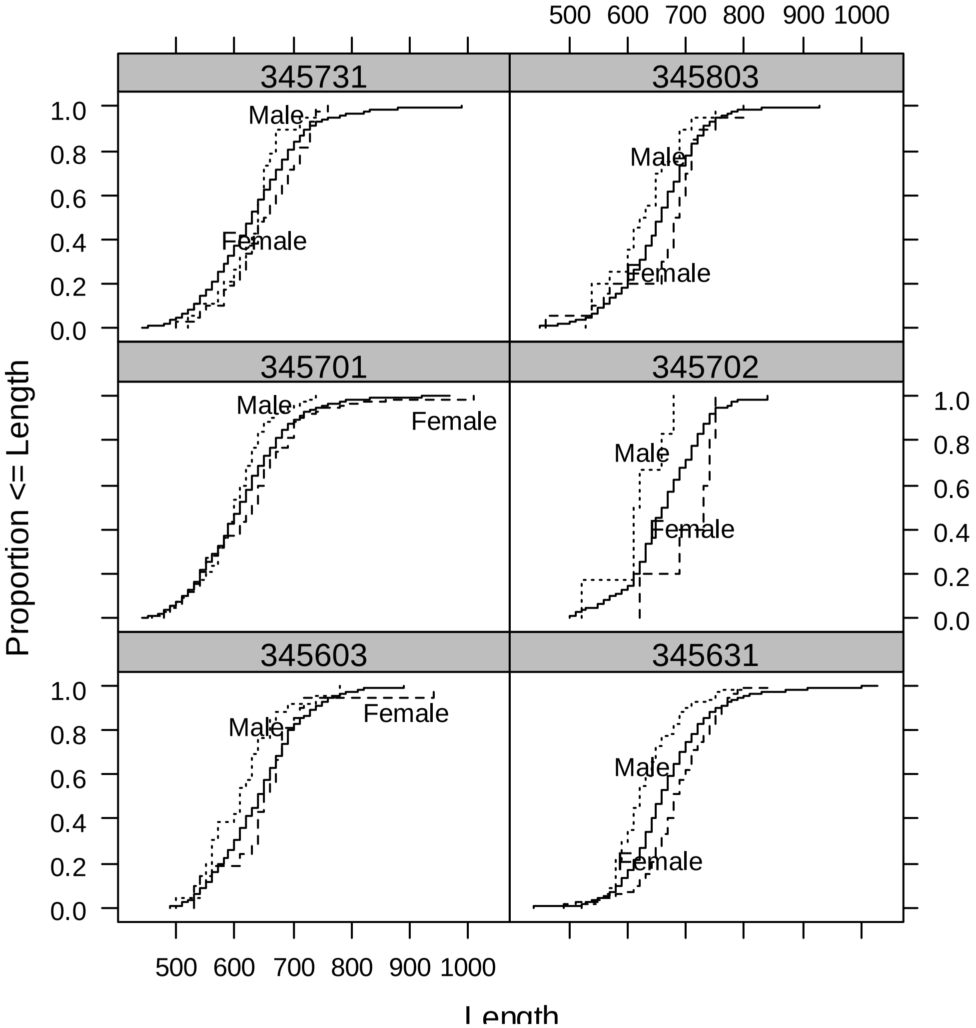


Figure 20.–Cumulative length-frequencies of unsexed (solid line) and of male (dotted line) and female (dashed line) sablefish sampled during pot survey by ADF&G statistical area.

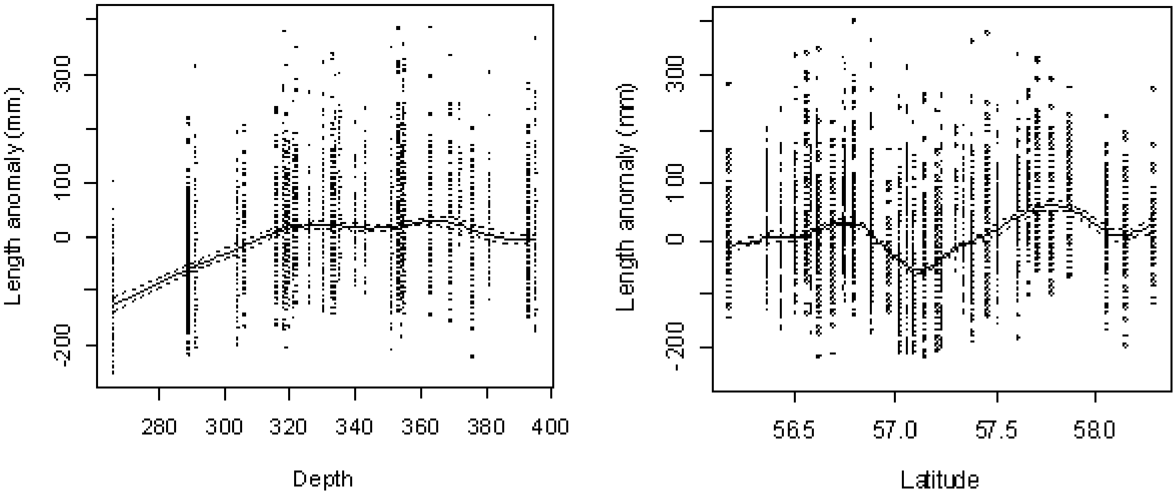


Figure 21.–Estimated trends in average length with depth and latitude during 2006 pot survey with 95% confidence bands based on GAM (amount of smoothing determined by cross-validation).

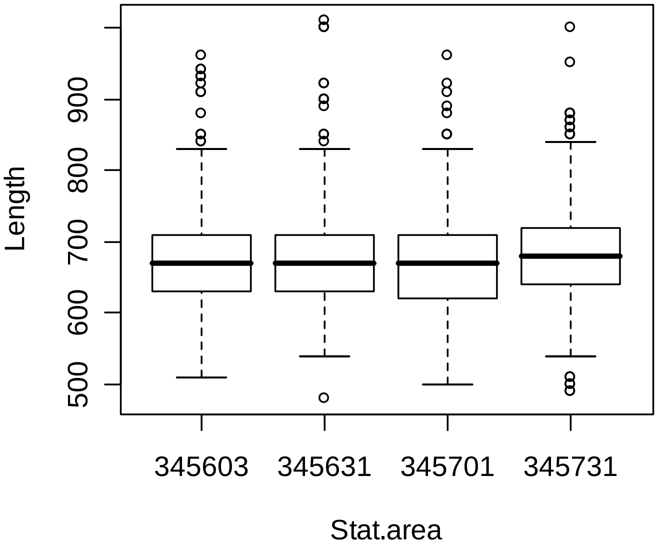


Figure 22.–Box plots of length distribution of all sablefish caught during longline survey by statistical area.

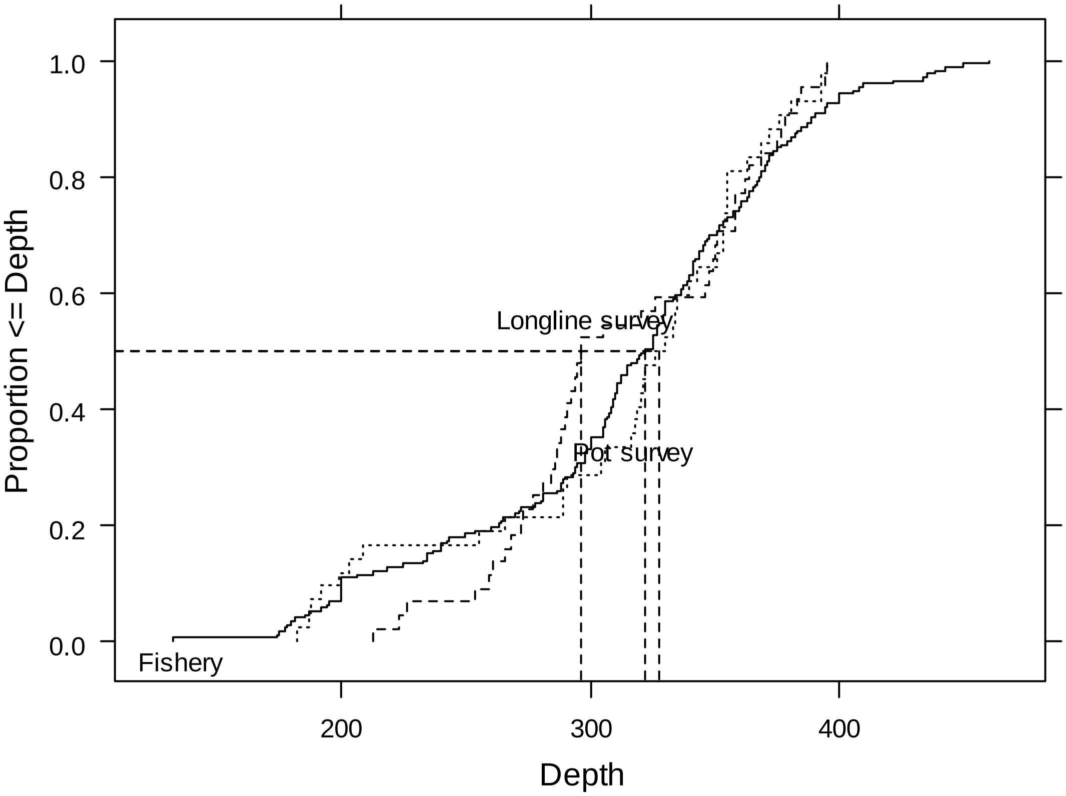


Figure 23.–Cumulative depth distribution of pot (dotted line) and longline survey stations (dashed line) and fishery trips (solid line) (n = 42, n = 44, and n = 173, respectively). Kolmogorov-Smirnov tests do not indicate significant differences, except a marginally significant difference between longline survey and fishery (D = 0.216, p = 0.075).

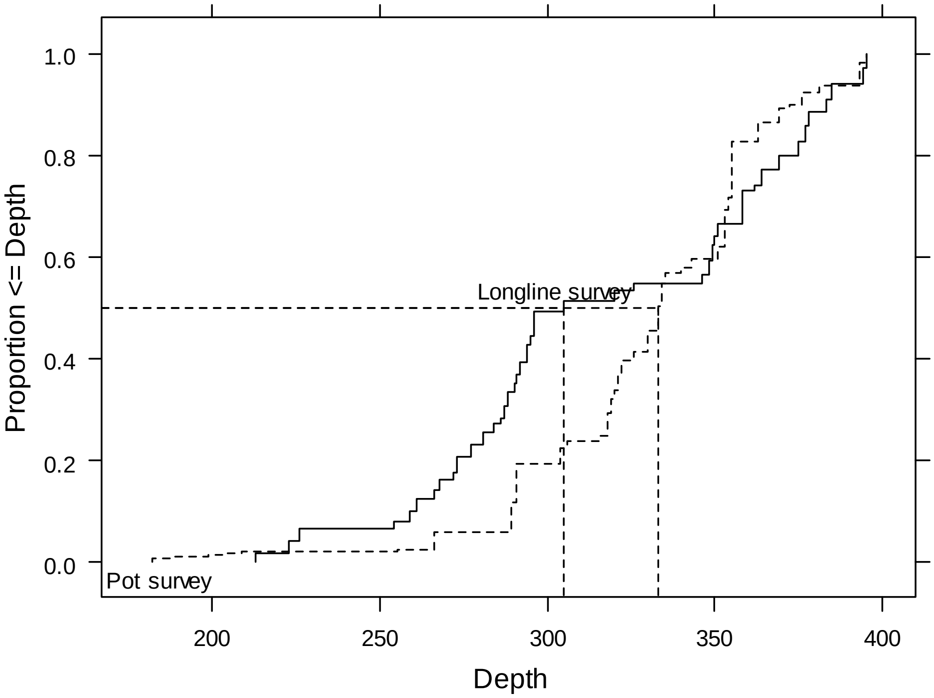


Figure 24.–Cumulative depth distribution of all sablefish caught in the pot (dashed line) and longline surveys (solid line) (n = 7259 and n =1482, respectively). Kolmogorov-Smirnov test is highly significant (D = 0.298, p < 0.0001).

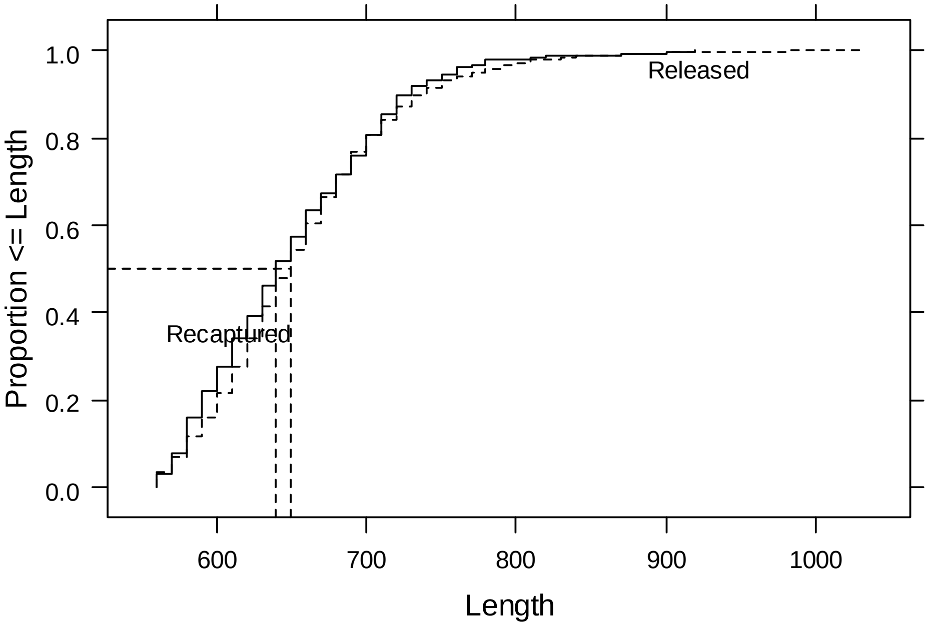


Figure 25.–Cumulative size distribution of tagged fish released in the survey (dotted line) and recovered in the fishery (solid line) (Kolmogorov-Smirnov test: D=0.063, p = 0.053).

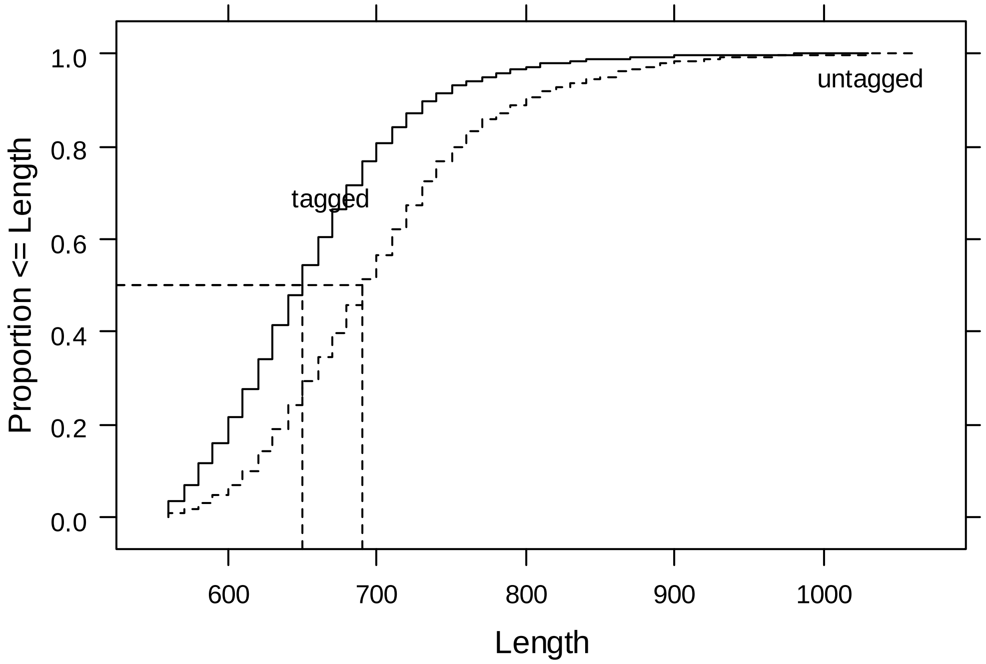


Figure 26.–Cumulative size distribution of tagged fish recovered in the fishery and untagged fish recovered in the fishery (only fish 560 mm or greater were considered. Kolmogorov-Smirnov test: D = 0.2662, p < 0.0001).

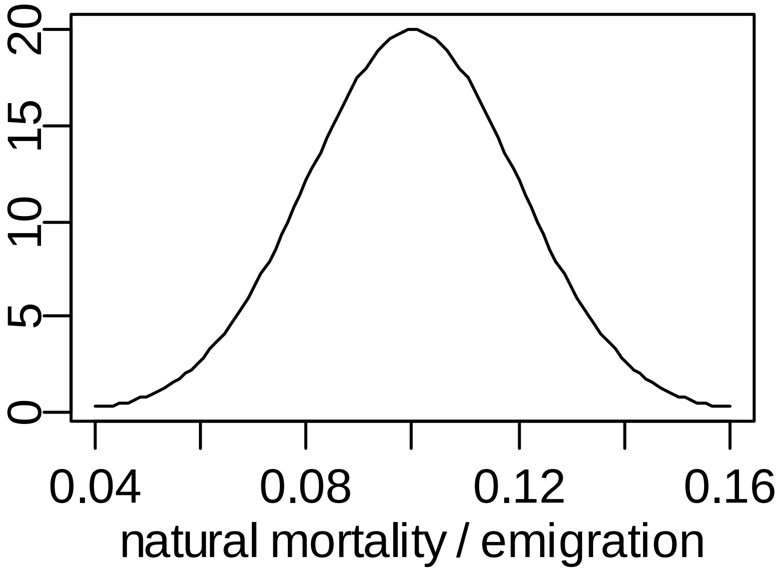


Figure 27.–Prior distribution on annual instantaneous natural mortality rate *m* used in models 2–4.

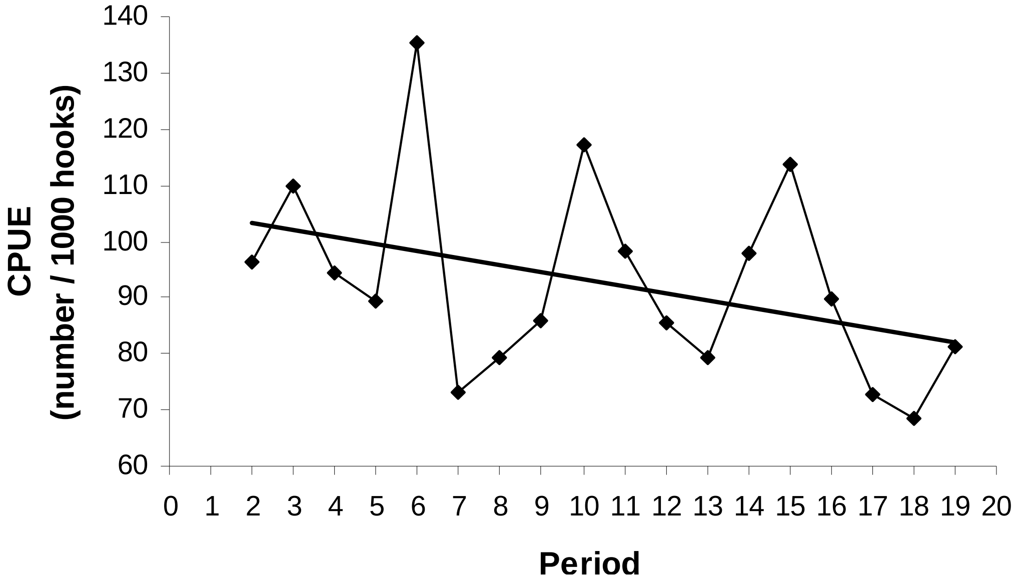


Figure 28–Fishery CPUE averaged across deliveries (trips) occurring within one of 18 consecutive 5-day periods with linear regression line (Linear regression weighted by the number of trips within each period: t = -1.178, p = 0.256. Number of trips contributing to a period's average was generally much lower in later periods, which therefore received lower weights in the regression).

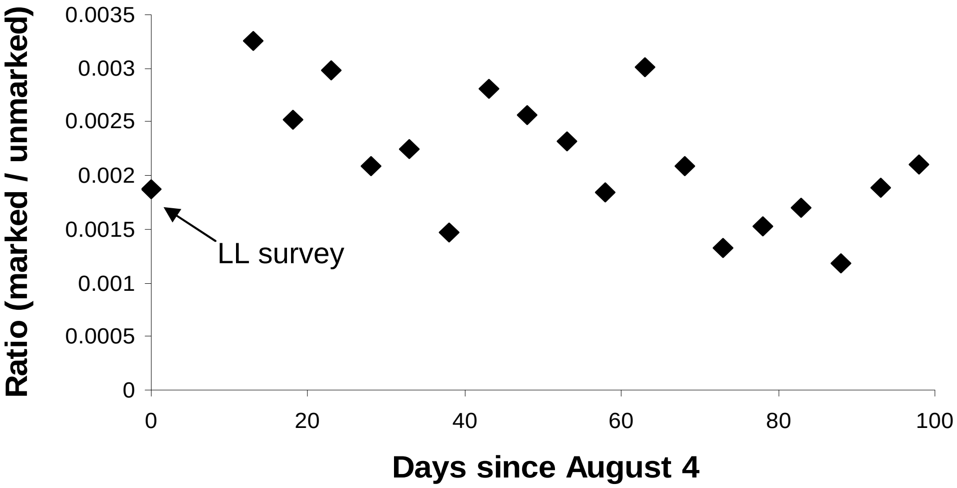


Figure 29–Trend in observed minus expected number of marks by sampling period (longline survey and 18 fishery periods, denoted by mid-point of 5-day periods) based on model B.

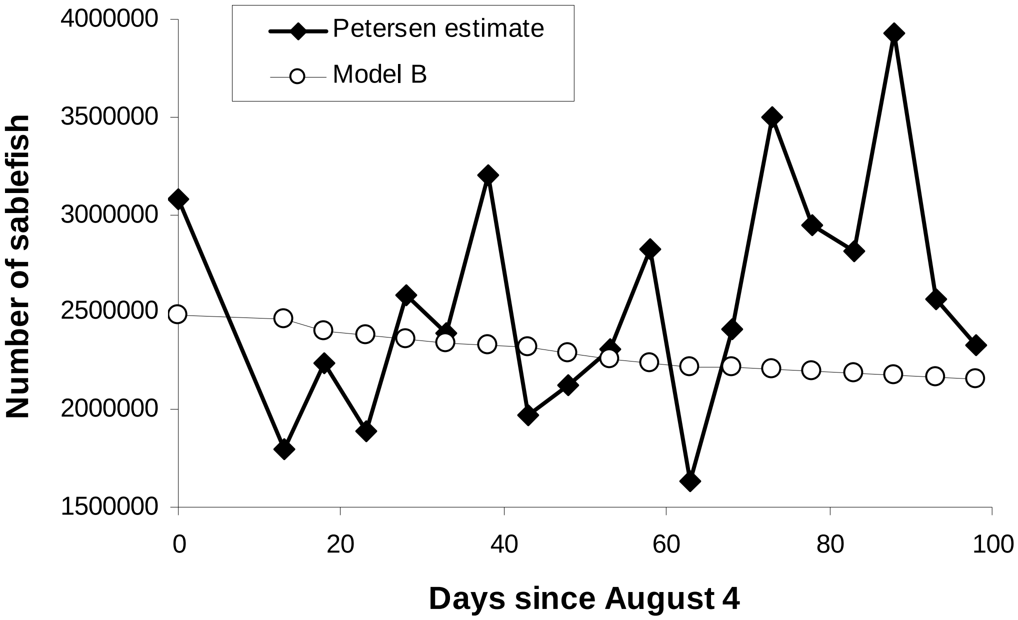


Figure 30.–Model B estimated trend in sablefish abundance and corresponding time-stratified Petersen estimates of abundance for longline survey and 18 consecutive 5-day fishery periods, plotted at mid-point of survey and fishing periods.

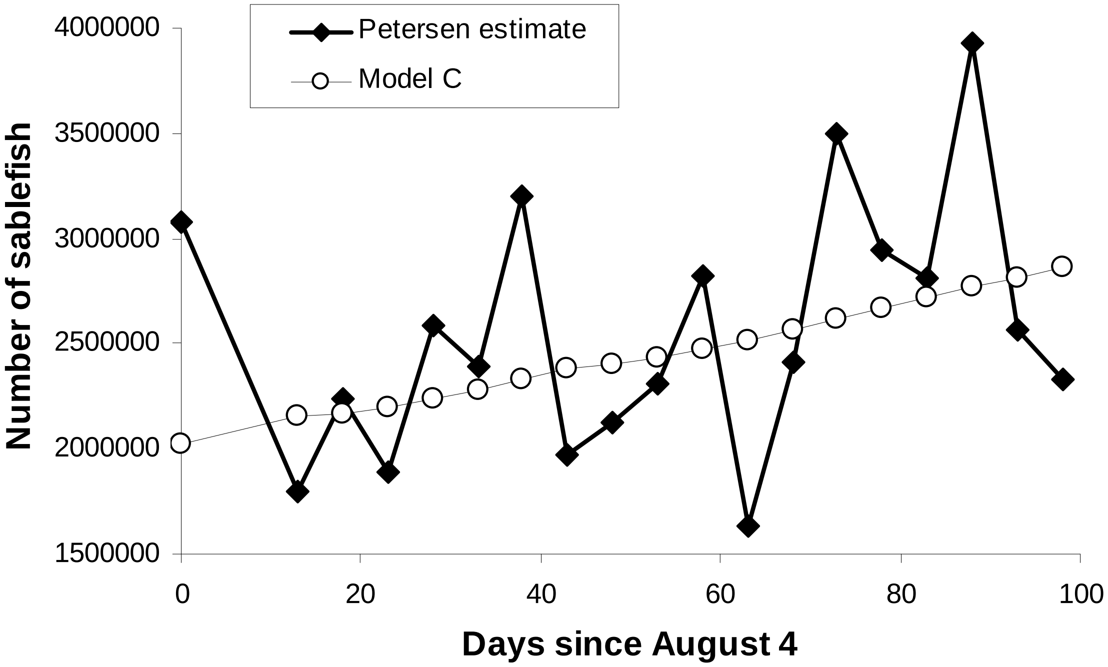


Figure 31.–Model C estimates of sablefish abundance and corresponding time-stratified Petersen estimates of abundance for longline survey and 18 consecutive 5-day fishery periods, plotted at mid-point of survey and fishing periods.

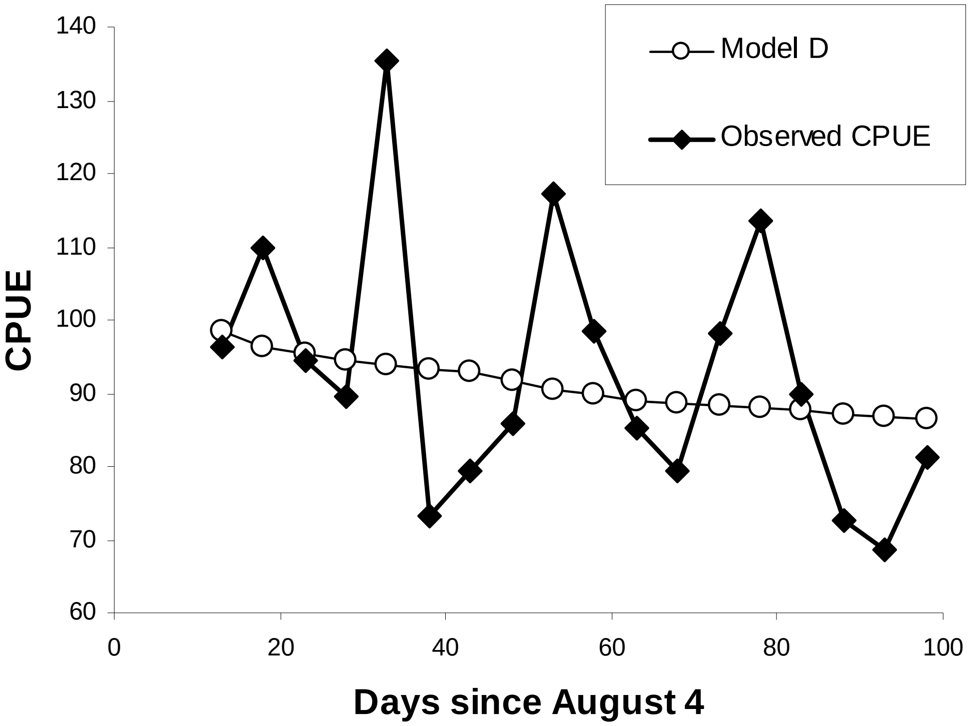


Figure 32.–Model D estimated trend in fishery CPUE and observed fishery CPUEs for 18 consecutive 5-day fishery periods, plotted at mid-points of periods.

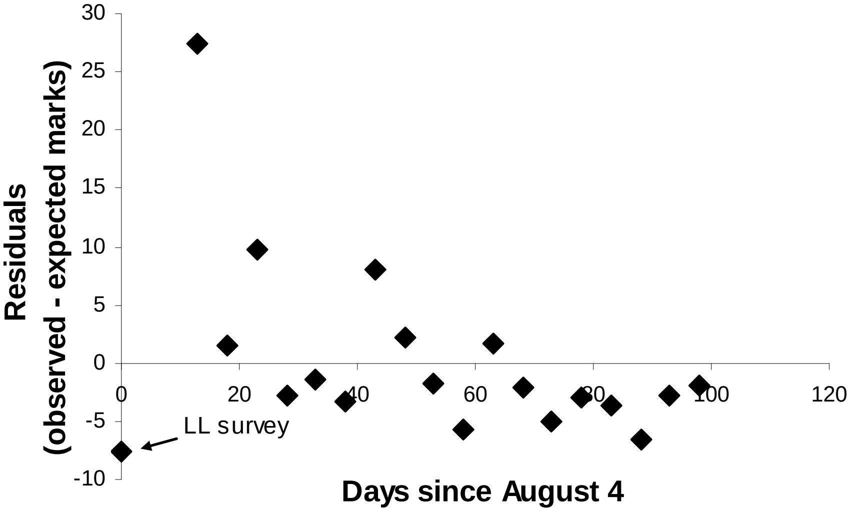


Figure 33.–Trend in observed minus expected number of marks by sampling period (LL survey and 18 5-day fishery periods, plotted at mid-points of periods) based on model D.

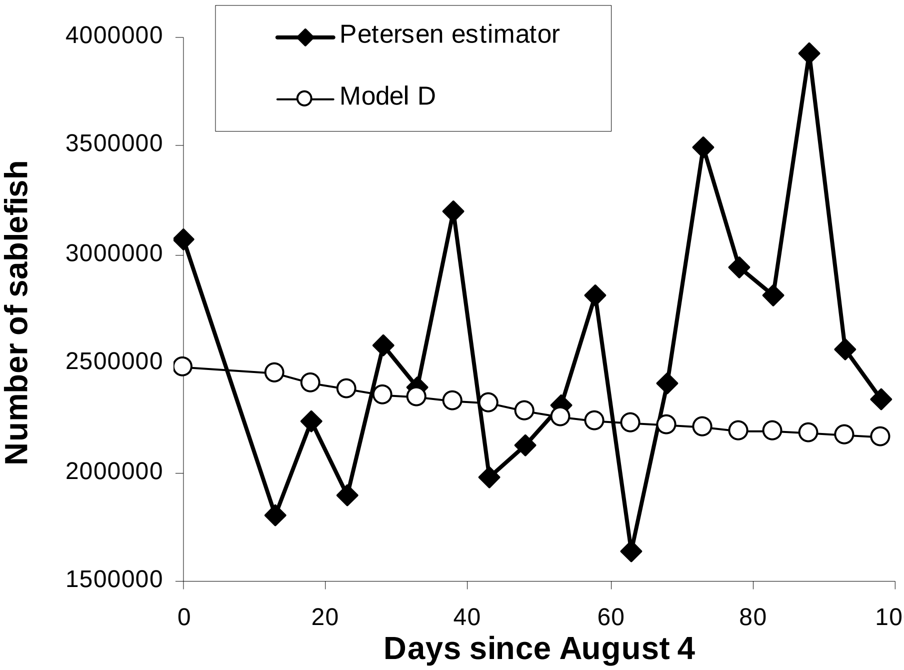


Figure 34.–Model D estimated trend in sablefish abundance over time and corresponding time-stratified Petersen estimates of abundance for longline survey and 18 consecutive 5-day fishery periods, plotted at mid-points of periods.

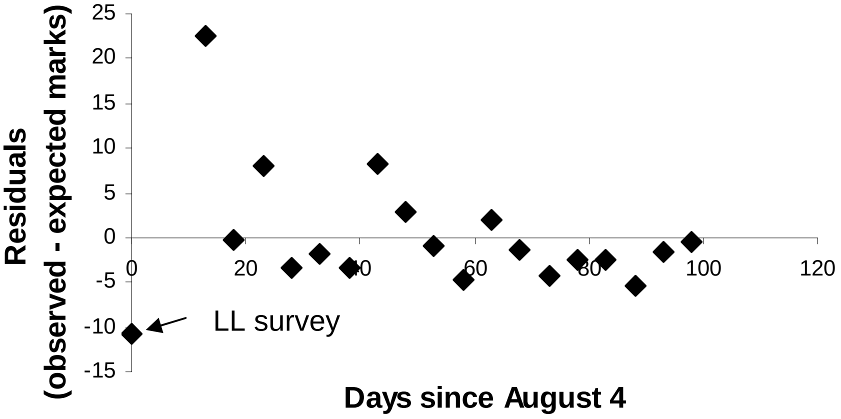


Figure 35.–Trend in observed minus expected number of marks by sampling period (LL survey and 18 5-day fishery periods, plotted at mid-points of periods) based on model E.

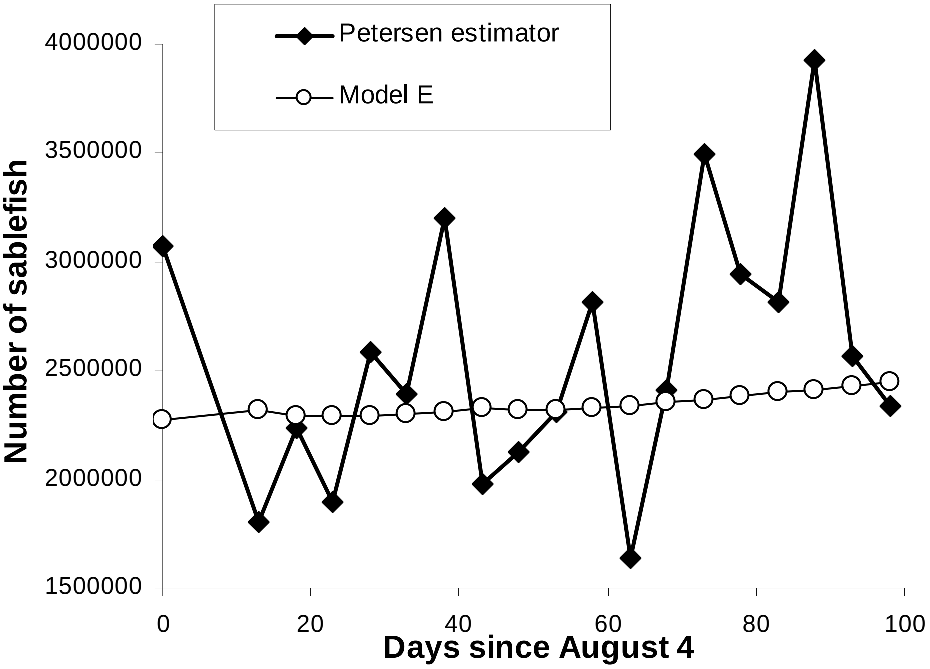


Figure 36.–Model E estimated trend in sablefish abundance over time and corresponding time-stratified Petersen estimates of abundance for longline survey and 18 consecutive 5-day fishery periods, plotted at mid-points of periods.

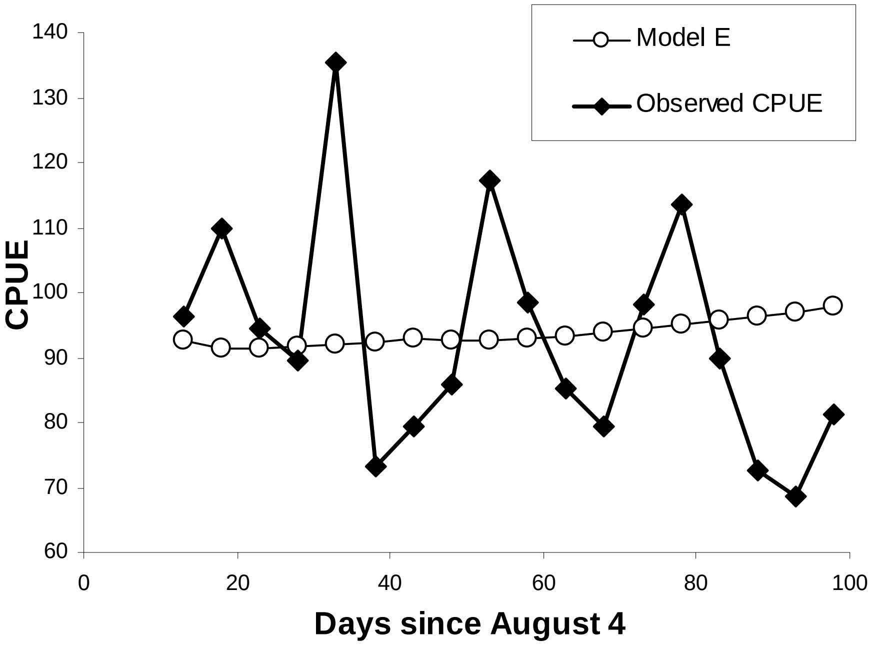


Figure 37.–Model E estimated trend in fishery CPUE and observed fishery CPUEs for 18 consecutive 5-day fishery periods, plotted at mid-points of periods.

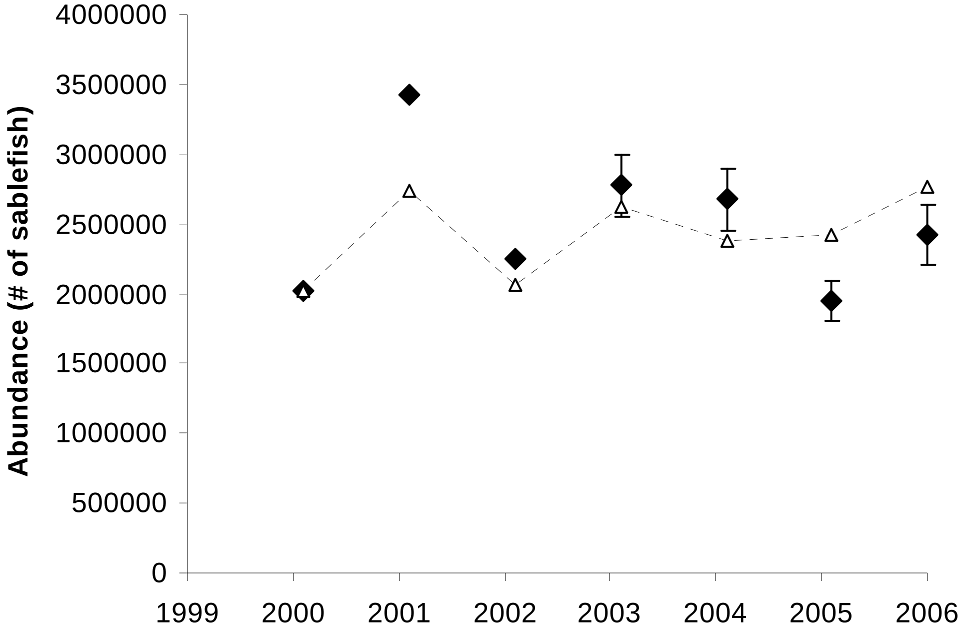


Figure 38.–Historical estimates of numerical abundance and 90% confidence intervals of sablefish in Chatham Strait, 2000–2006. Estimates in 2000–2002 were converted from biomass estimates based on exploitation rates to total numerical abundance using average individual weights from the fishery. Dashed line is a smoothed estimate based on a first-order autoregressive model of order 1.

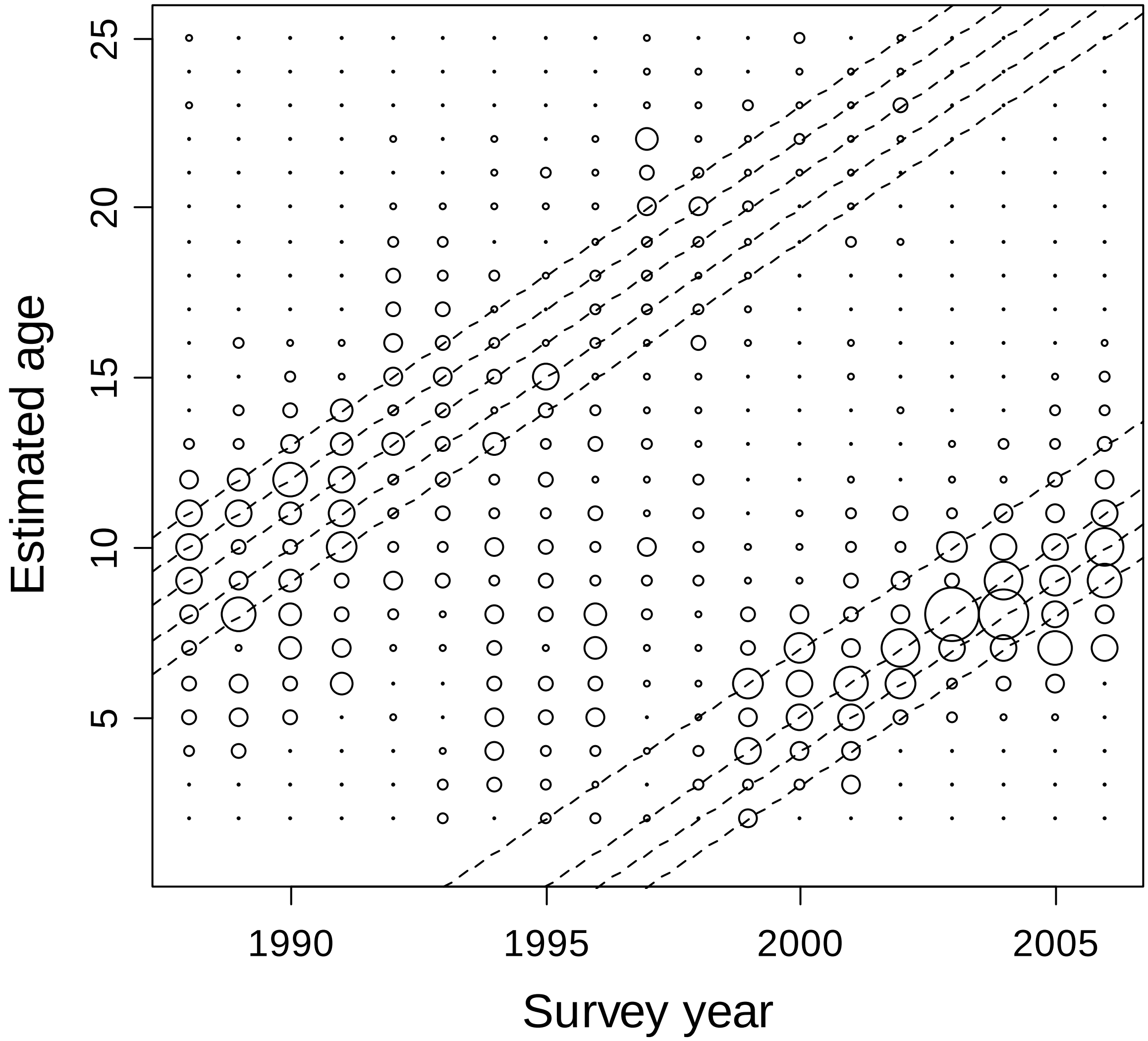


Figure 39.–Proportional age composition of sablefish caught in longline survey, 1988–2006. Circle size denotes proportion of age class within a given year. Strong year classes are denoted by dashed lines.

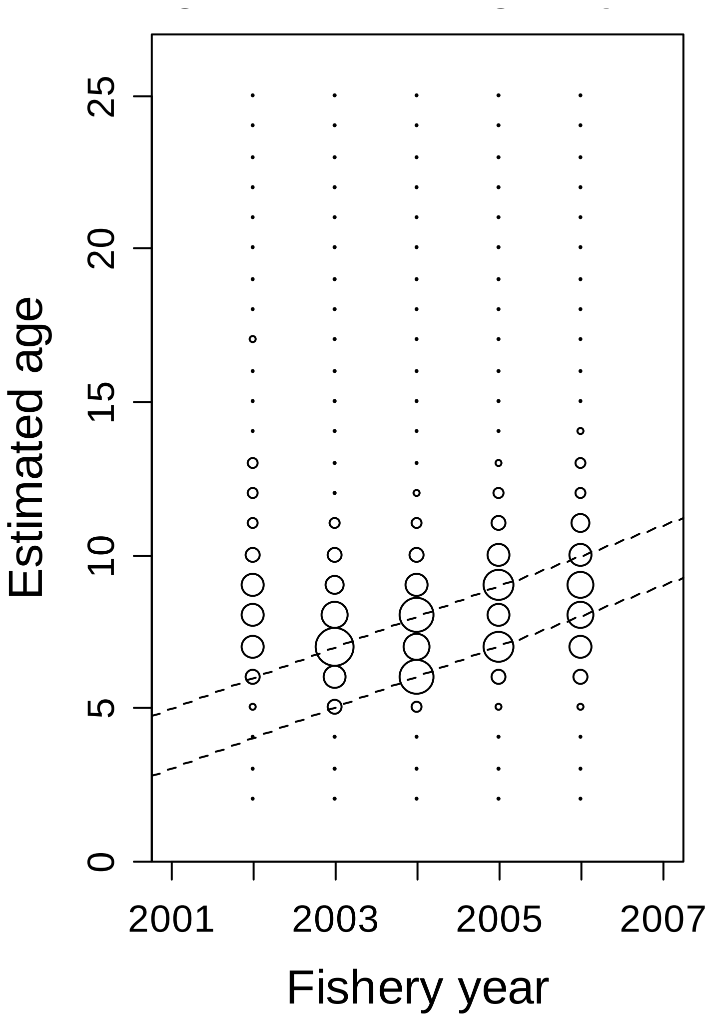


Figure 40.–Proportional age composition of sablefish caught in fishery, 2002–2006. Circle size denotes proportion of age class within a given year. Strong year classes are denoted by dashed lines.

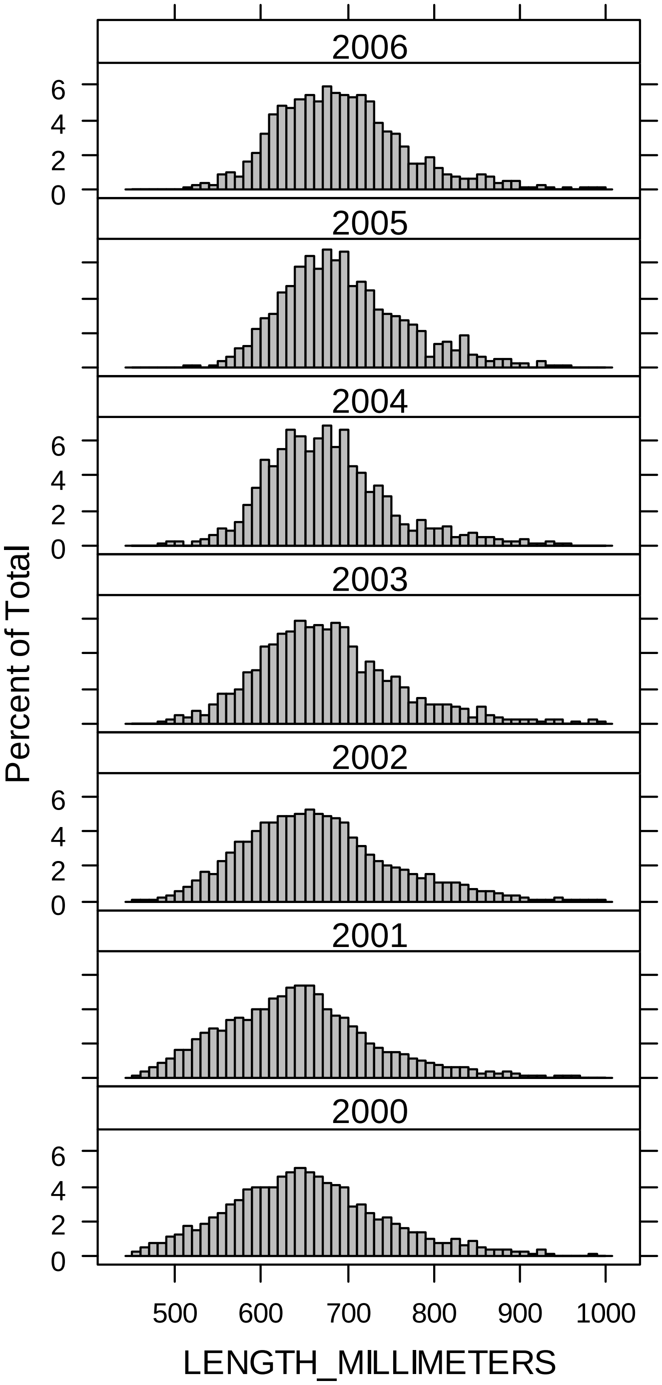


Figure 41.–Size composition in fishery catches, 2000–2006.

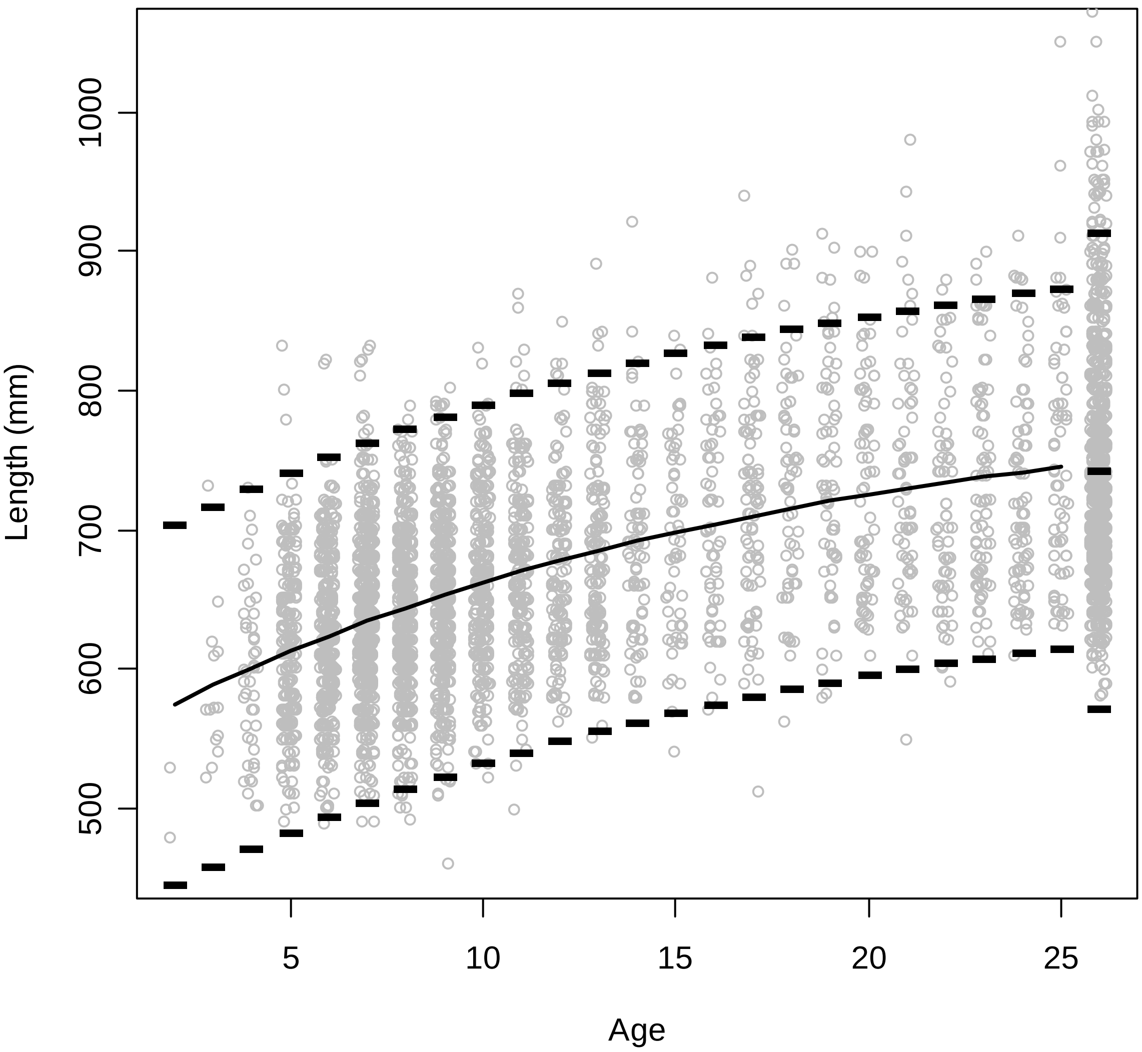


Figure 42.–Length at age for male and female sablefish sampled during 2002/2003 fishery with von Bertalanffy growth curve (solid line) for ages 2–25 and estimated means +/- 2 standard deviations at each age (horizontal bars). Mean +/- 2 standard deviations at age 26 was estimated from pooled data for all ages 26 and older.

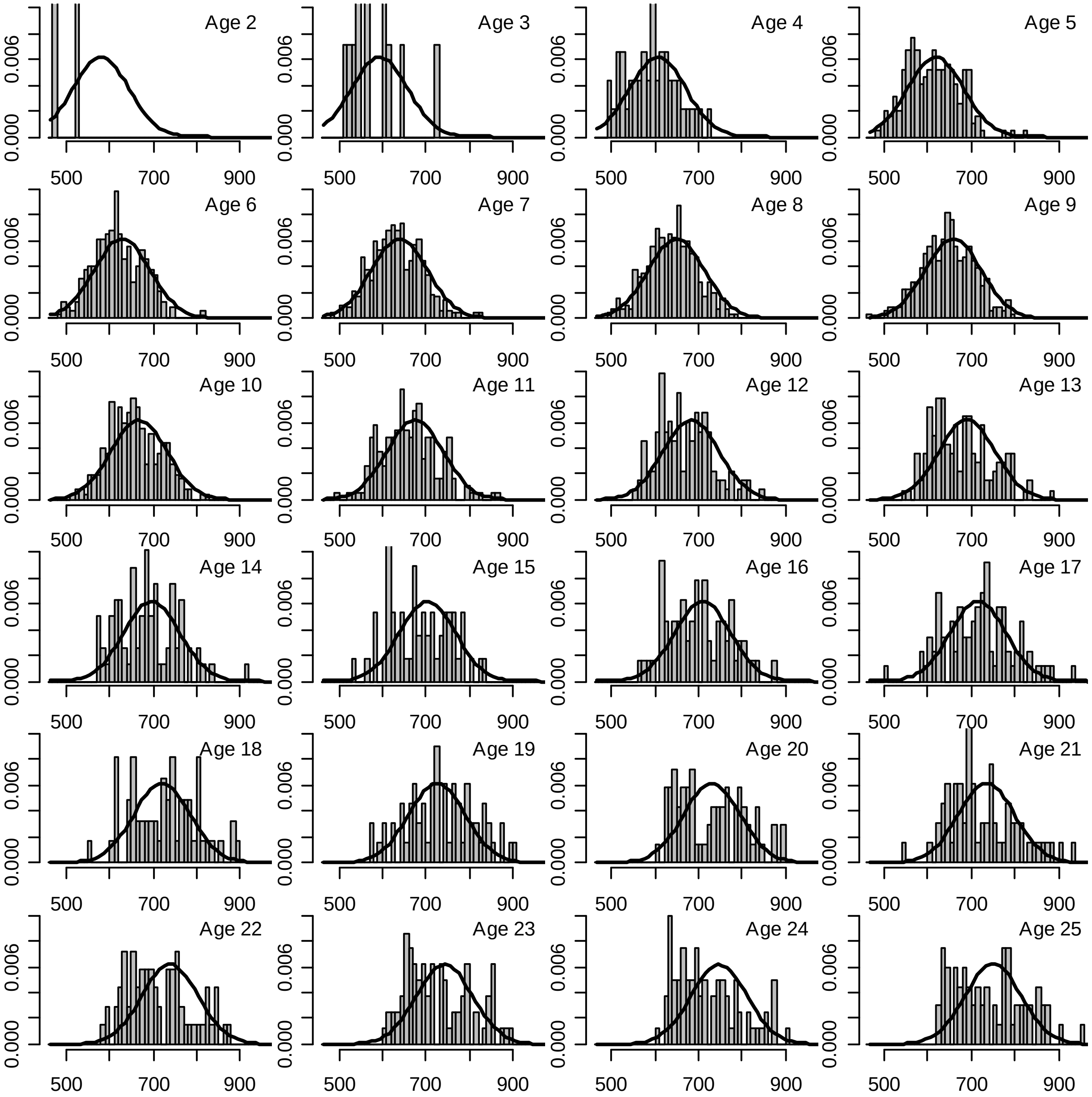


Figure 43.–Observed (bars) and estimated (lines) length distributions at ages 2–25 based on von Bertalanffy growth model with fitted normal distribution.

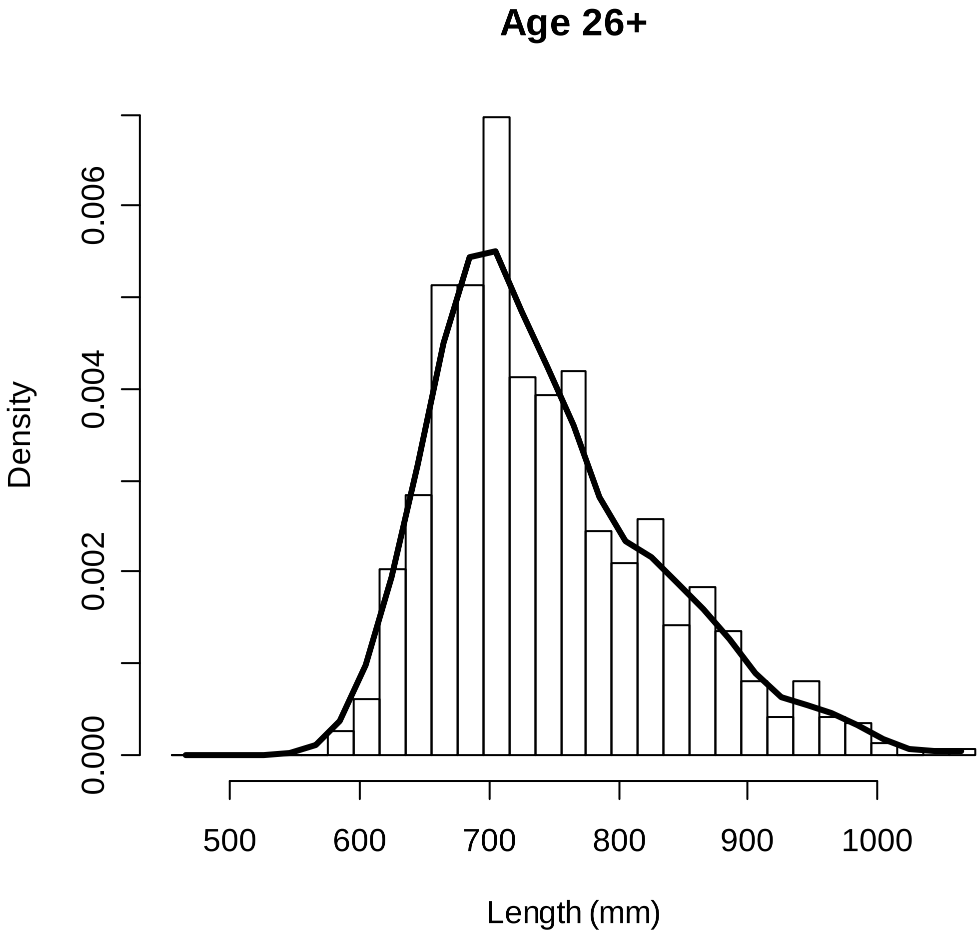


Figure 44.–Histogram of length distribution in the sablefish age 26+ group in 2002/03 with kernel density estimate, which was used to estimate proportion by length class.

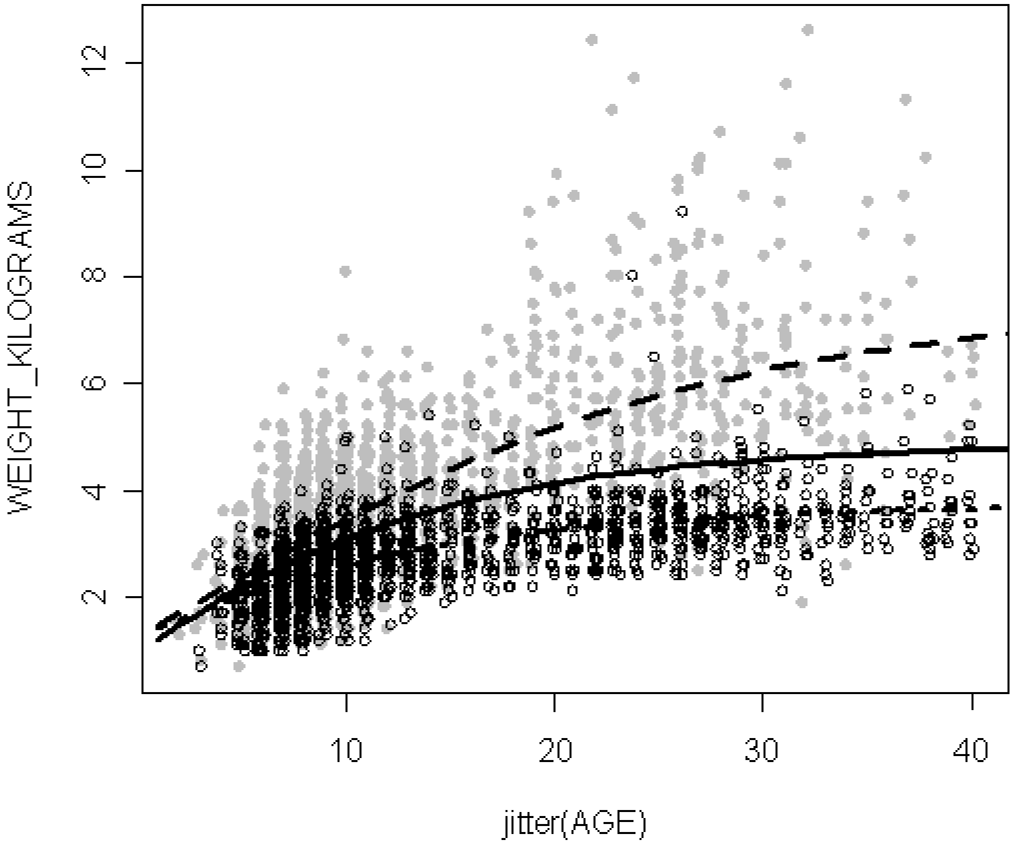


Figure 45.–Weight at age for male (black) and female (grey) sablefish caught during Chatham Strait longline survey with fitted von Bertalanffy growth models by sex (female: solid grey symbols, dashed line; males: circles, dotted line) and for both sexes combined (solid line).

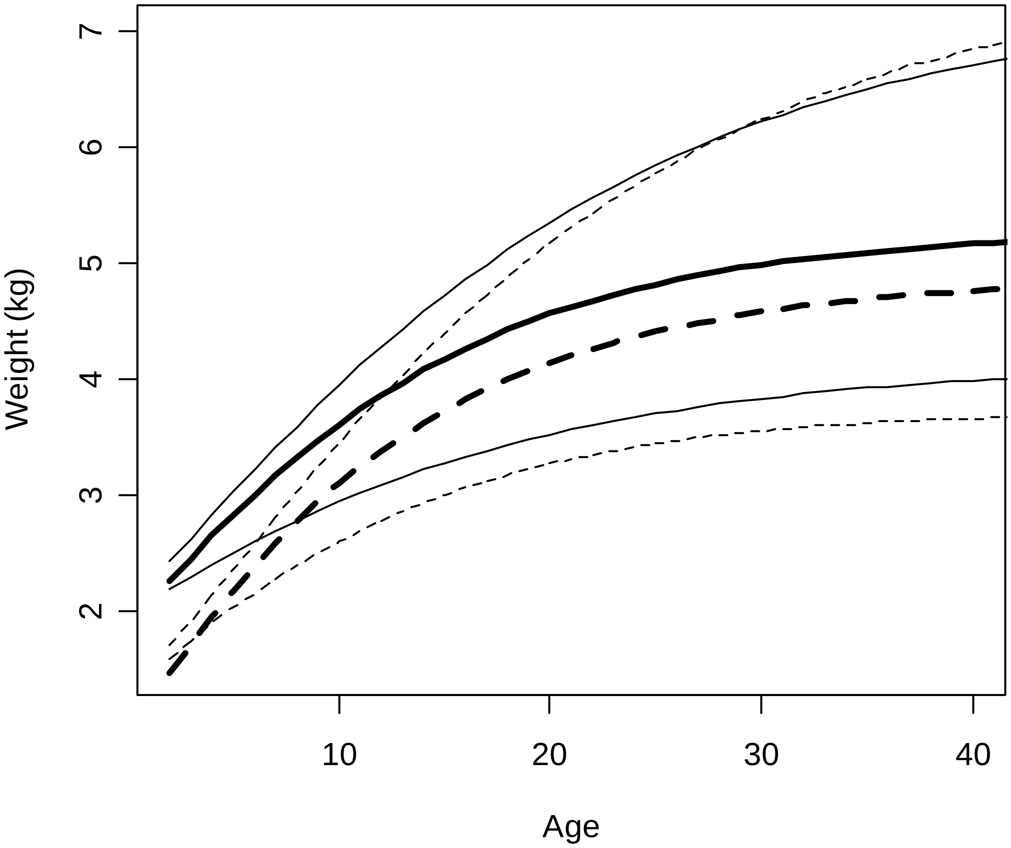


Figure 46.–Comparison of estimated mean weight at age between fishery-caught (solid lines) and survey-caught (dashed lines) sablefish: fitted von Bertalanffy models for females (upper lines), males (lower lines) and both sexes combined (heavy lines).

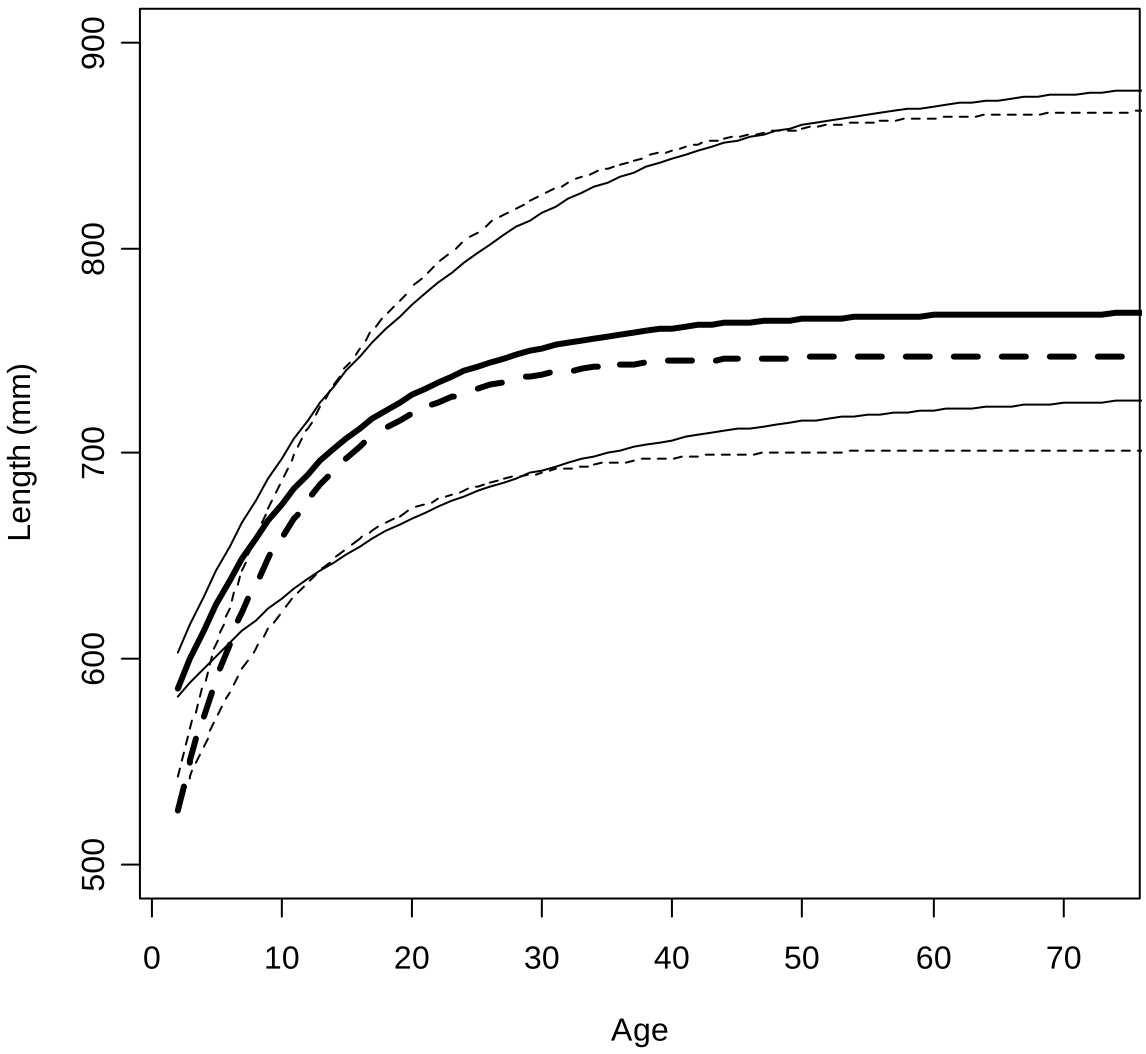


Figure 47.–Comparison of estimated mean length at age between fishery-caught (solid lines) and survey-caught (dashed lines) sablefish: fitted von Bertalanffy models for females (upper lines), males (lower lines) and both sexes combined (heavy lines).

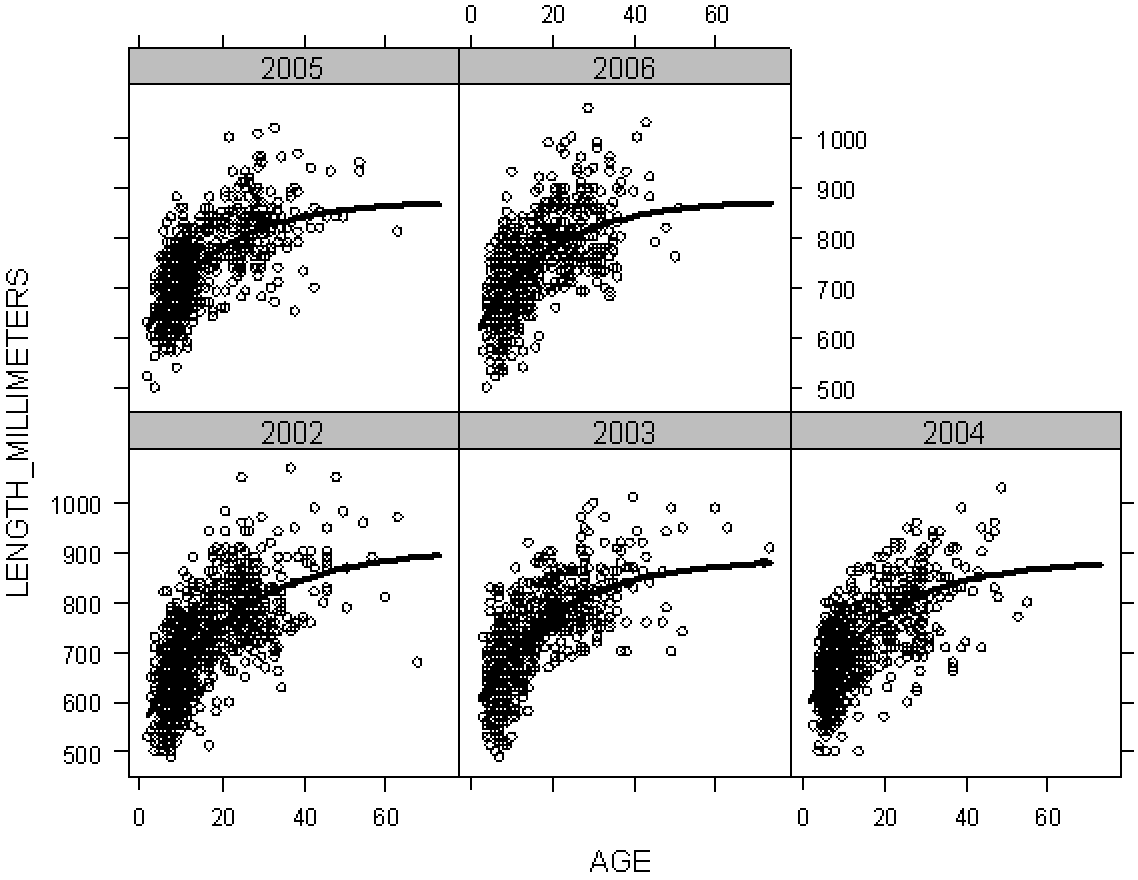


Figure 48.–Von Bertalanffy fits for female sablefish caught in the fishery by year based on mixed-effects model.

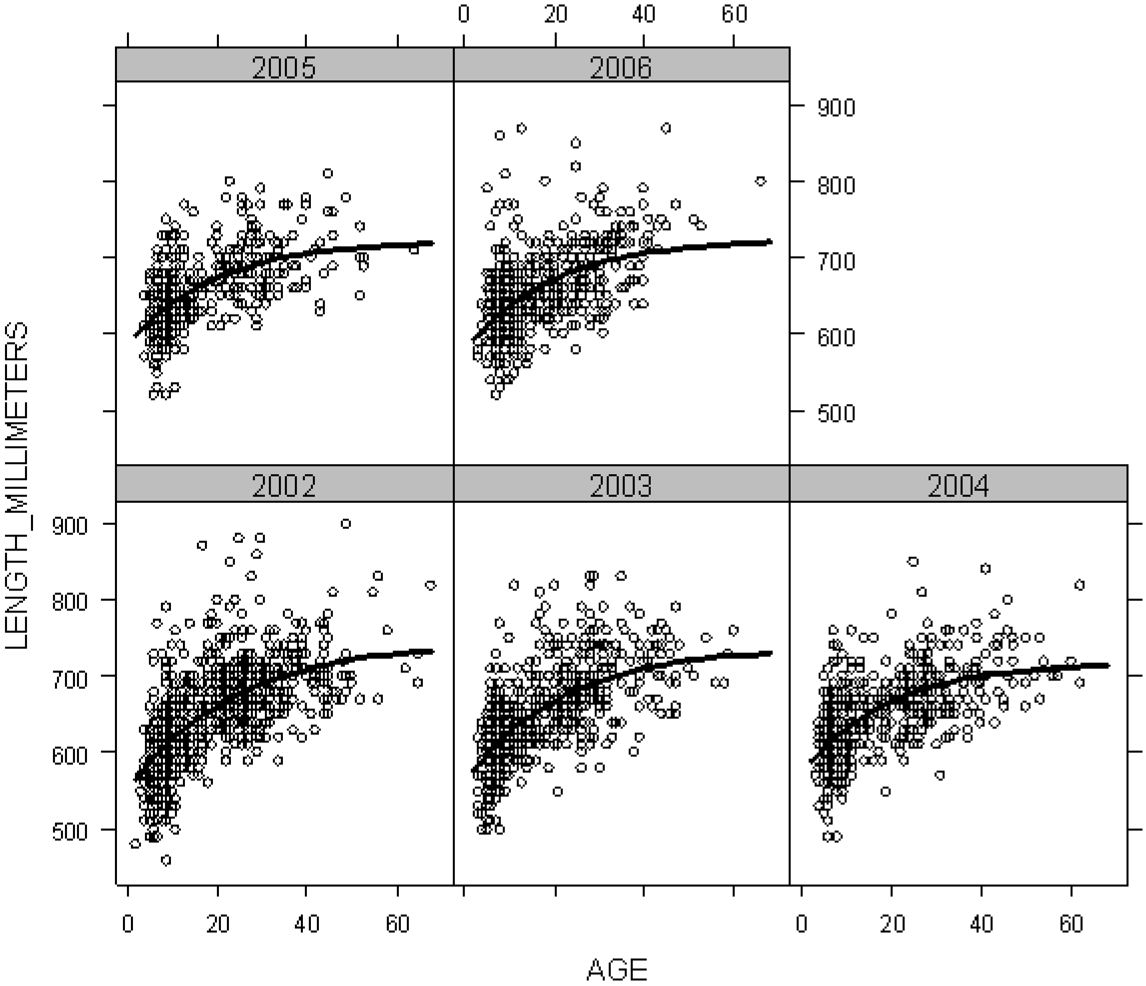
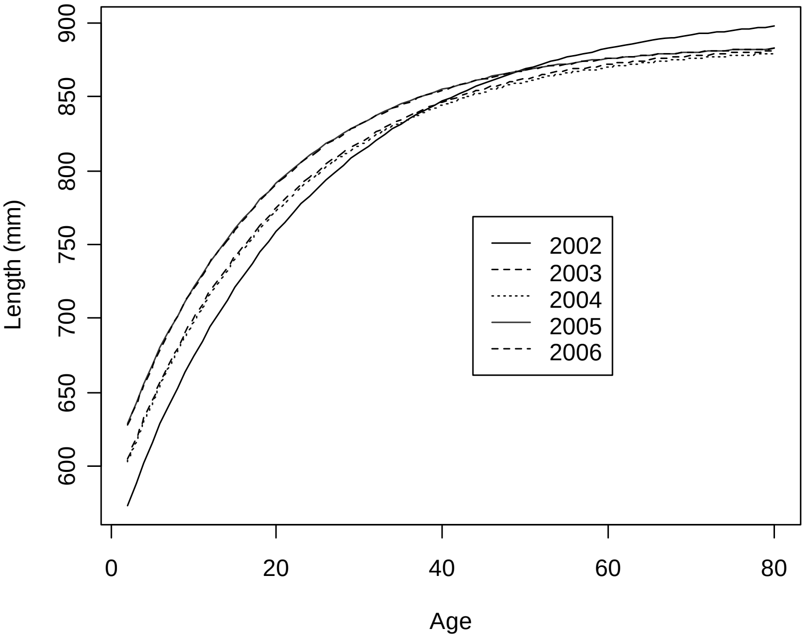


Figure 49.–Von Bertalanffy fits for male sablefish caught in the fishery by year based on mixed-effects model.



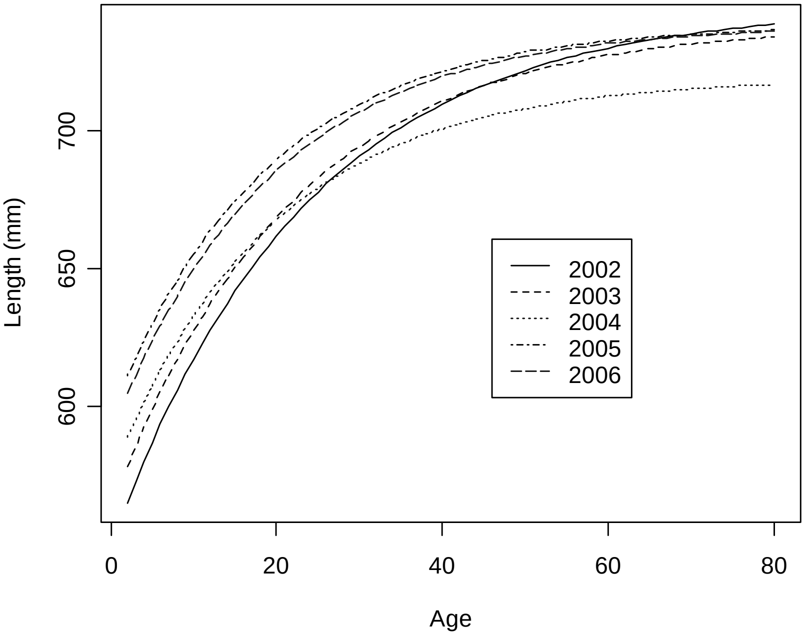


Figure 50.–Comparison of estimated von Bertalanffy growth curves for female (top) and male (bottom) sablefish among years, 2002–2006.

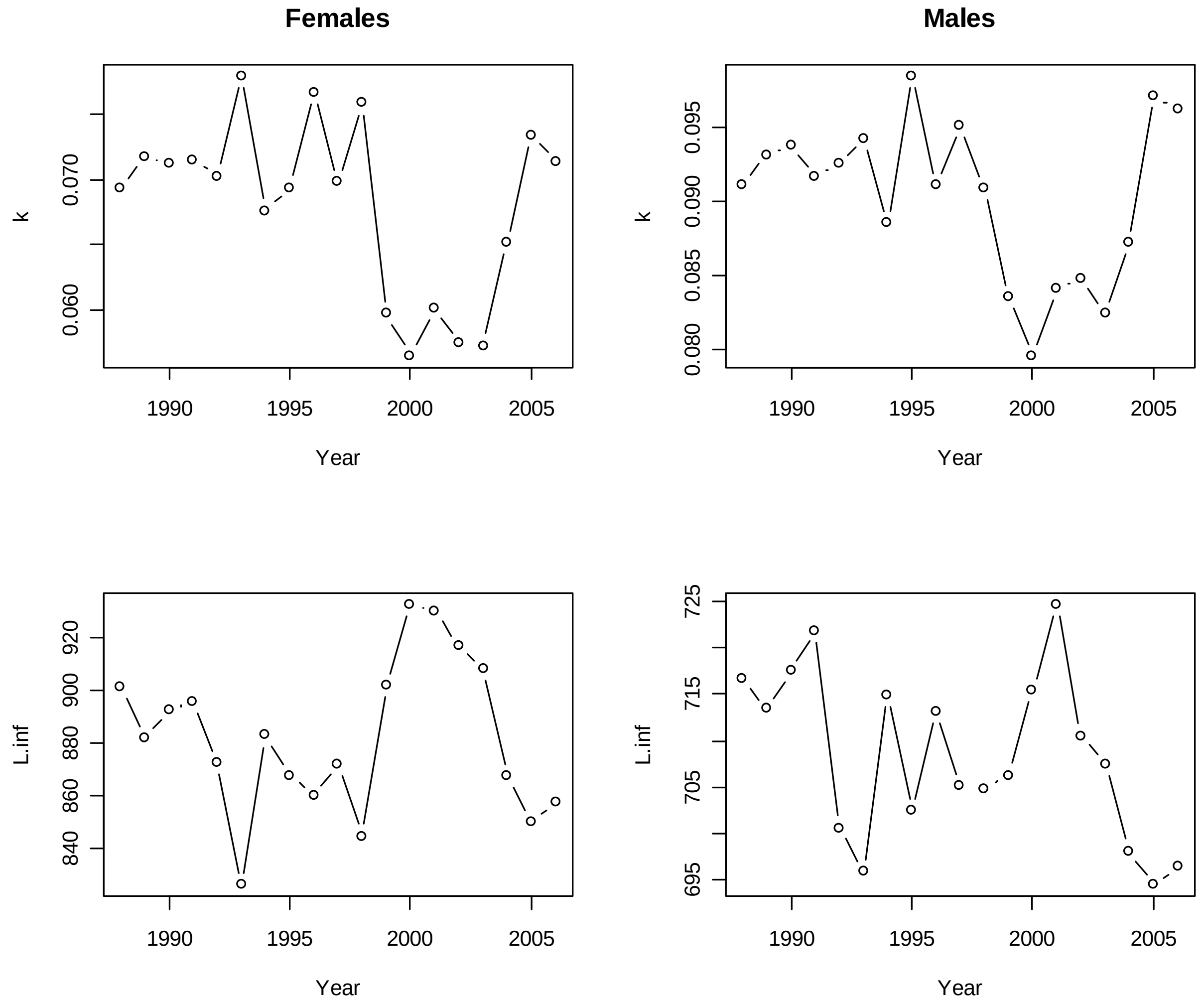


Figure 51.–Estimated growth parameters k (top) and L_∞ (bottom) for von Bertalanffy growth model fit to male and female sablefish sampled during longline survey by year (using a mixed-effects model).

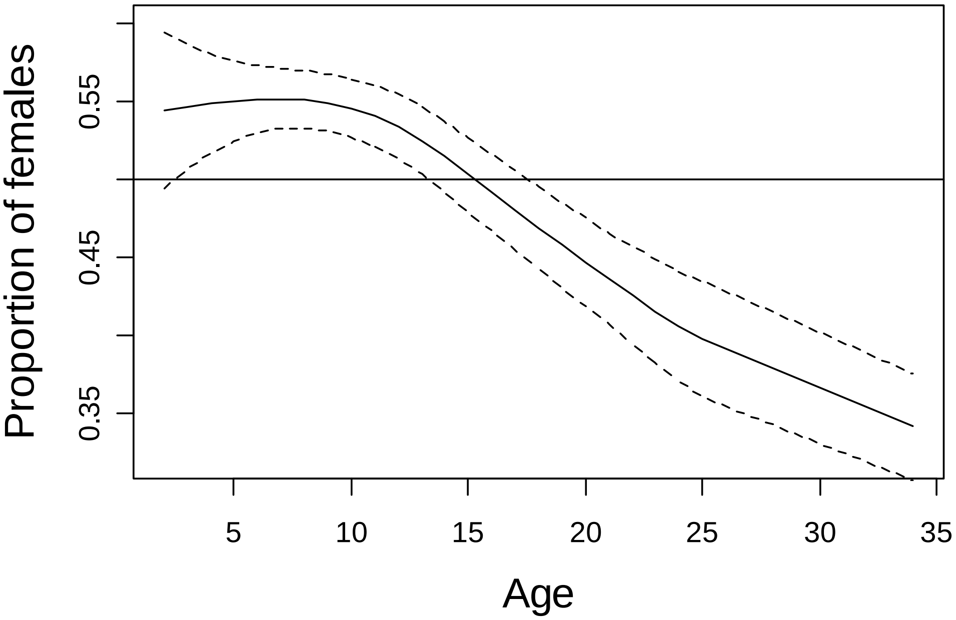


Figure 52.–Estimated proportion of females from the longline survey (1988–2006) as a smooth function of age, estimated via logistic (binomial) regression on age. Fish 26 years and older were combined and assigned to their average age (34 years). The horizontal line indicates equal proportion of males and females (0.5).

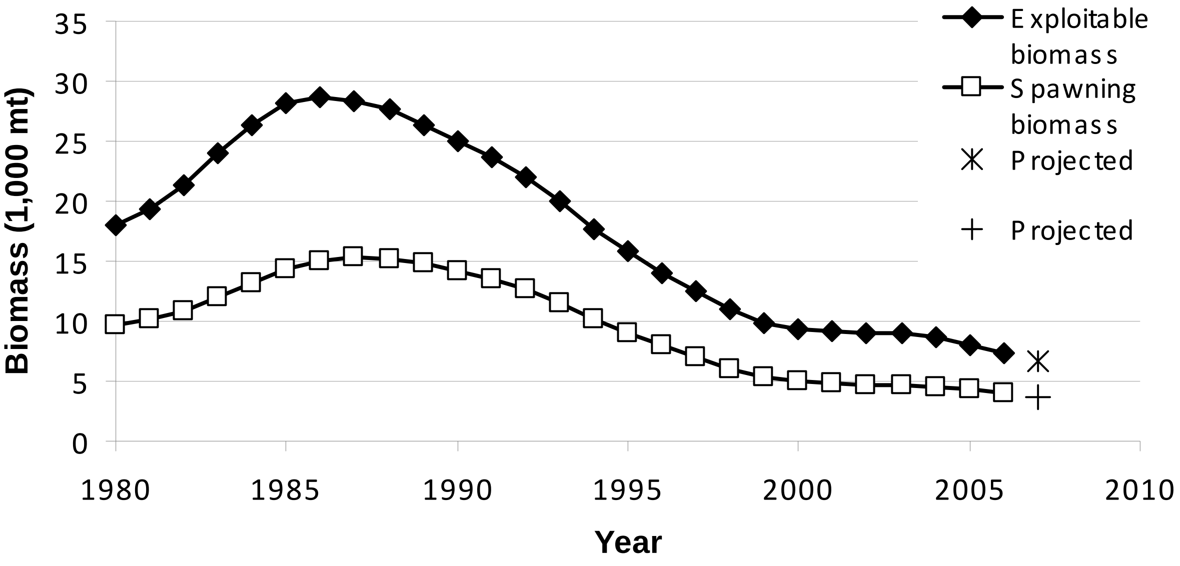


Figure 53.–Estimated spawning biomass and total exploitable biomass of sablefish in Chatham Strait, 1980–2006, under base model ($M = 0.1$, q for mark-recapture estimates fixed at 1).

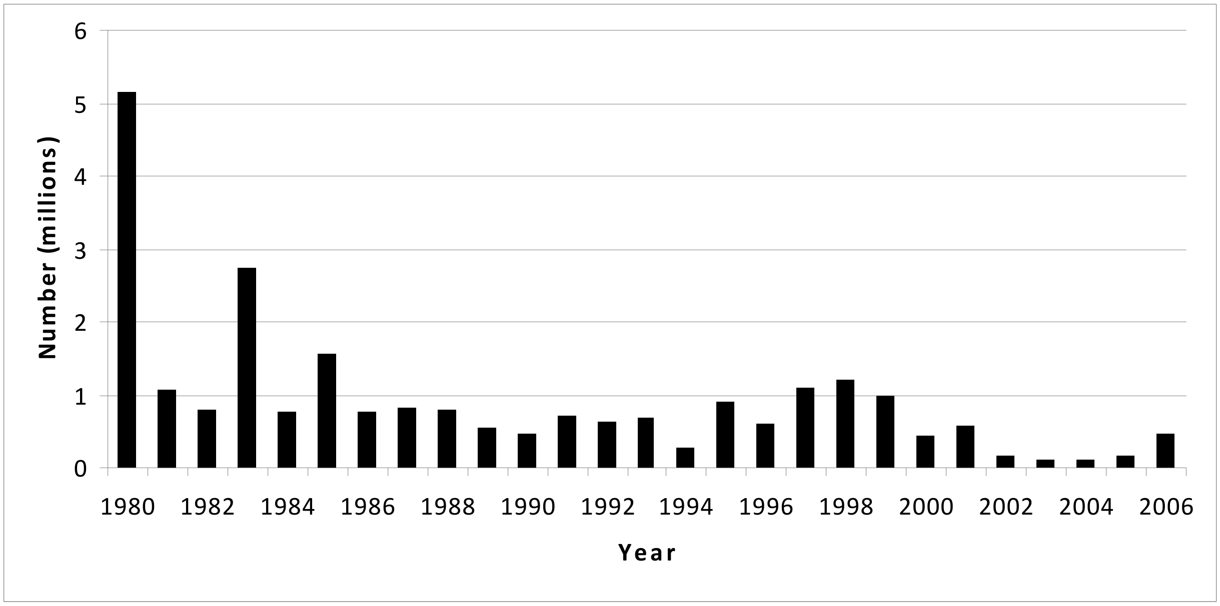


Figure 54.–Estimated recruitment trend of sablefish in Chatham Strait, 1980–2006, under base model ($M = 0.1$, q for mark-recapture estimates fixed at 1).

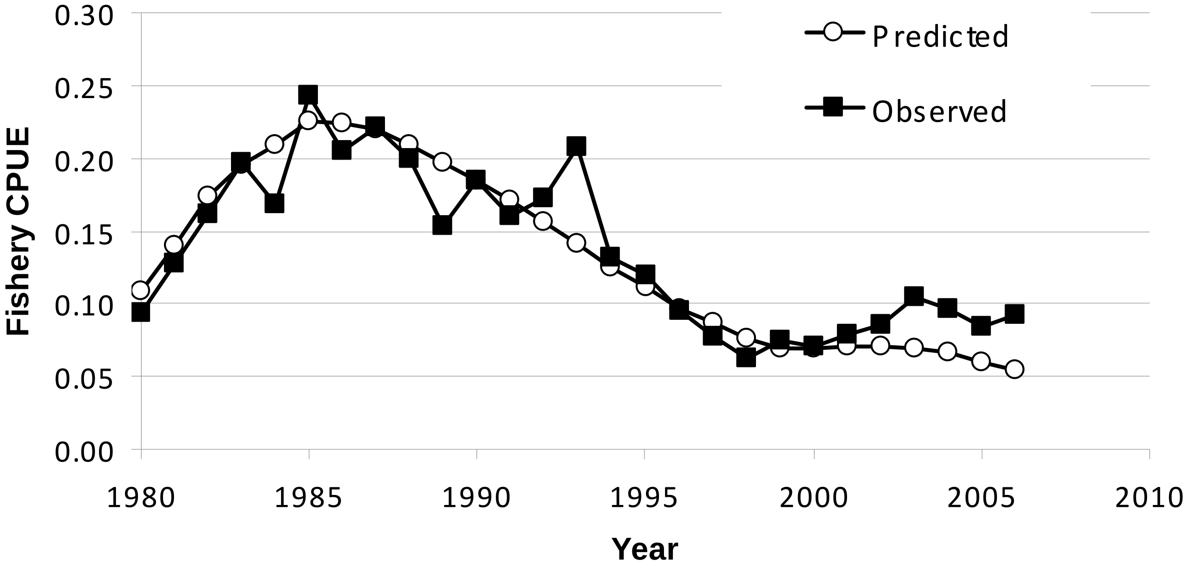


Figure 55.–Observed and model estimated fishery CPUE, 1980–2006, for Chatham Strait sablefish under base model ($M = 0.1$, q for mark-recapture estimates fixed at 1).

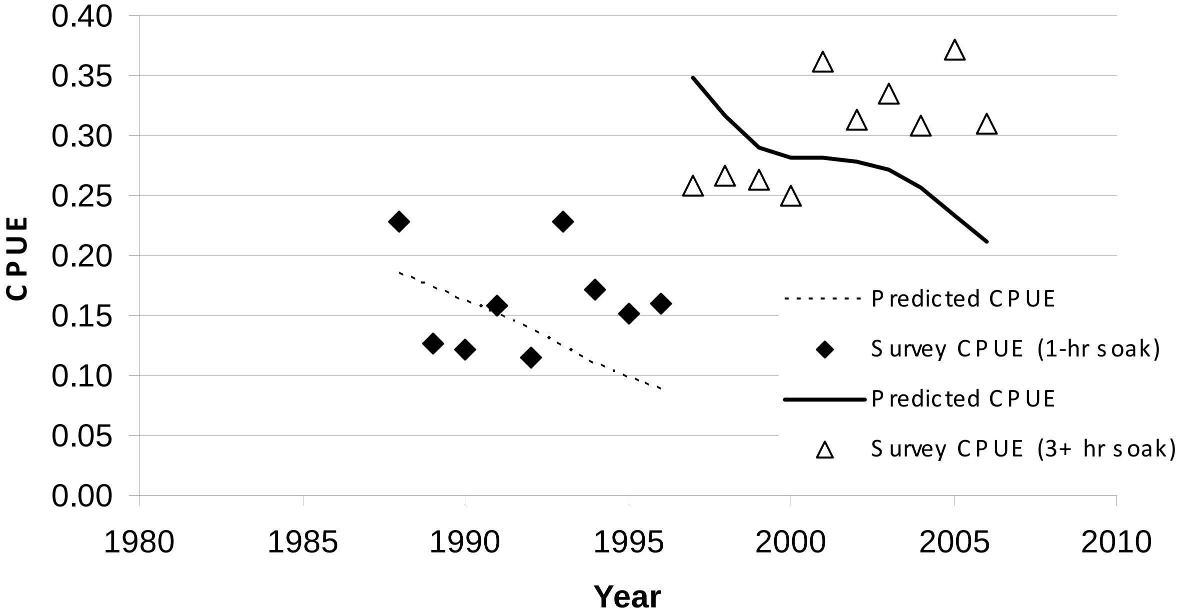


Figure 56.–Observed and model estimated longline survey CPUEs, 1988–1996 and 1997–2006, for Chatham Strait sablefish under base model ($M = 0.1$, q for mark-recapture estimates fixed at 1).

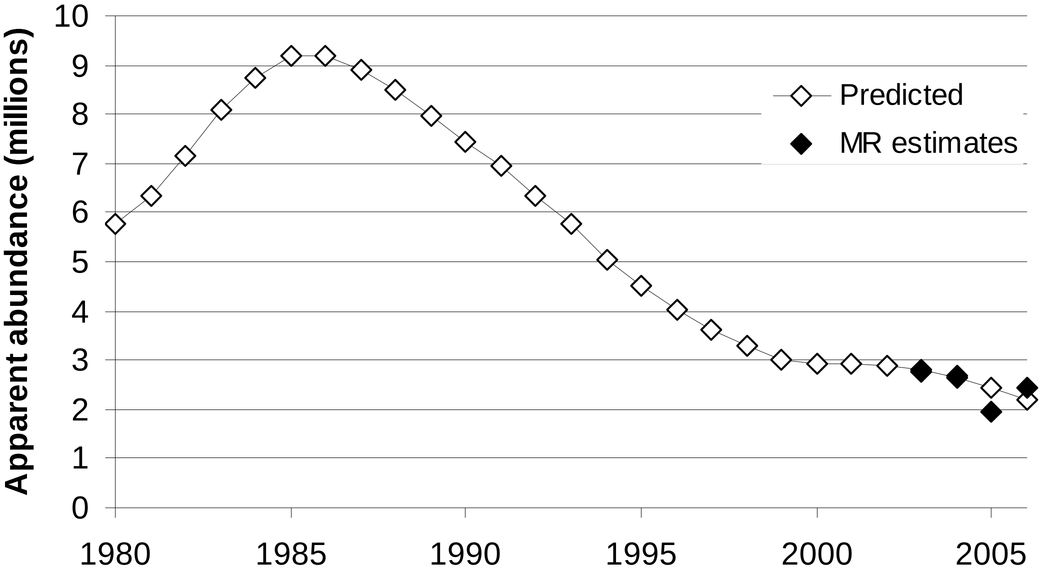


Figure 57.–Mark-recapture estimates of abundance, 2003–2006, and predicted abundance of sablefish at time of longline survey, 1980–2006, under base model (*M* = 0.1, *q* for mark-recapture estimates fixed at 1).

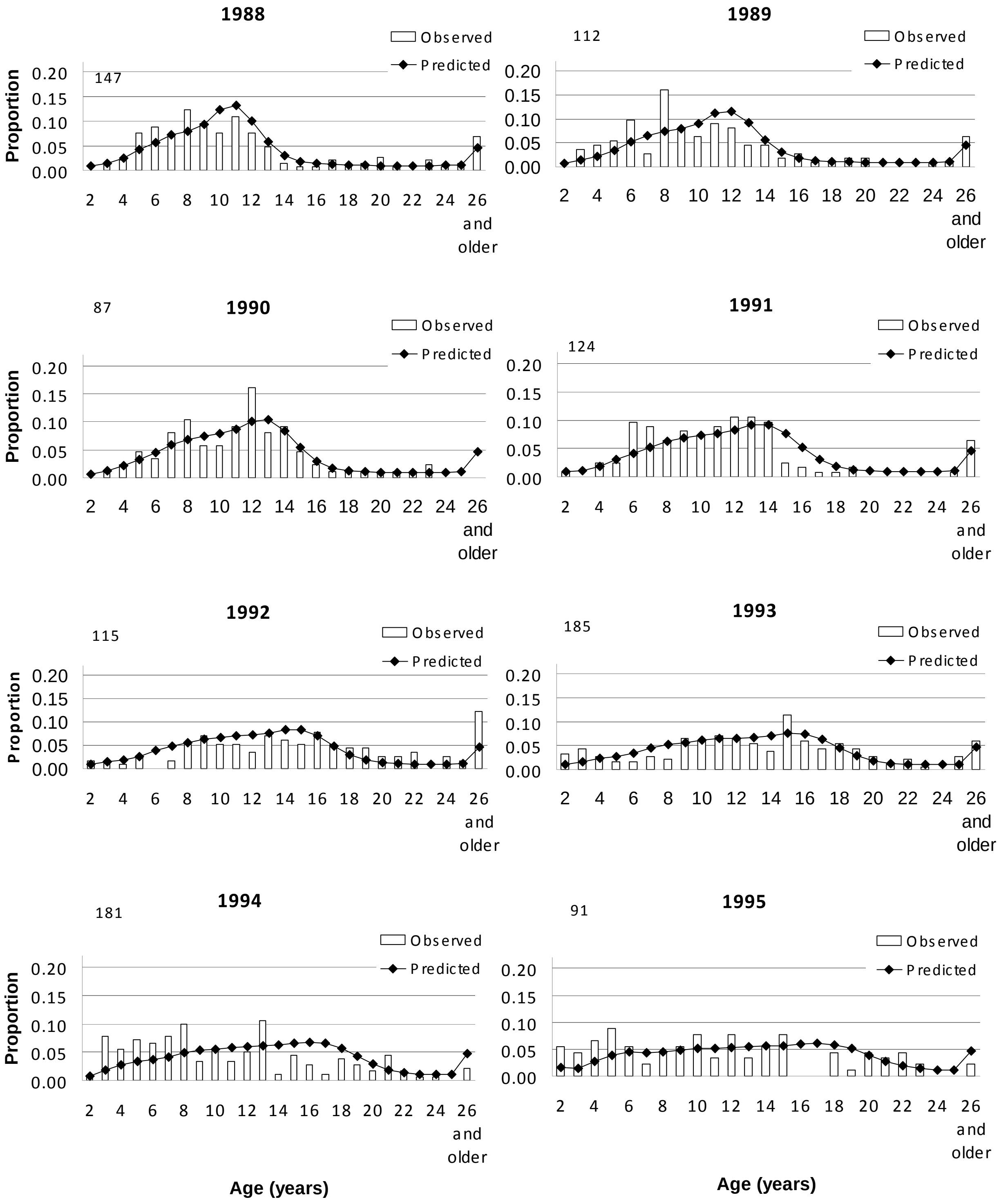


Figure 58.–Observed and predicted age composition during longline survey, 1988–2006, under base model (*M* = 0.1, *q* for mark-recapture estimates fixed at 1). Numbers in the upper left corner of each figure denote sample size, i.e., number of fish that were aged in a given year.

-continued-

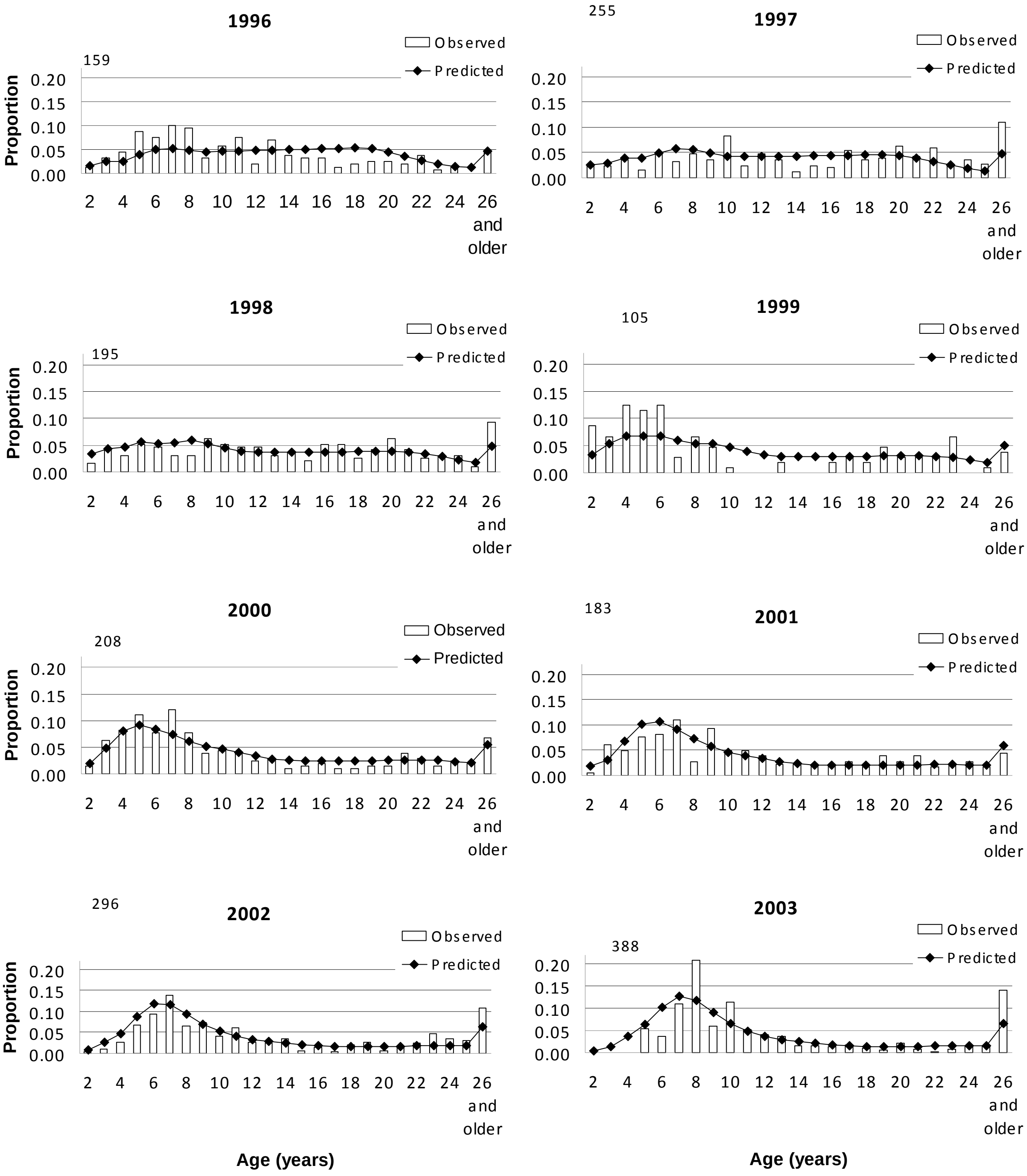


Figure 58.–Page 2 of 3.

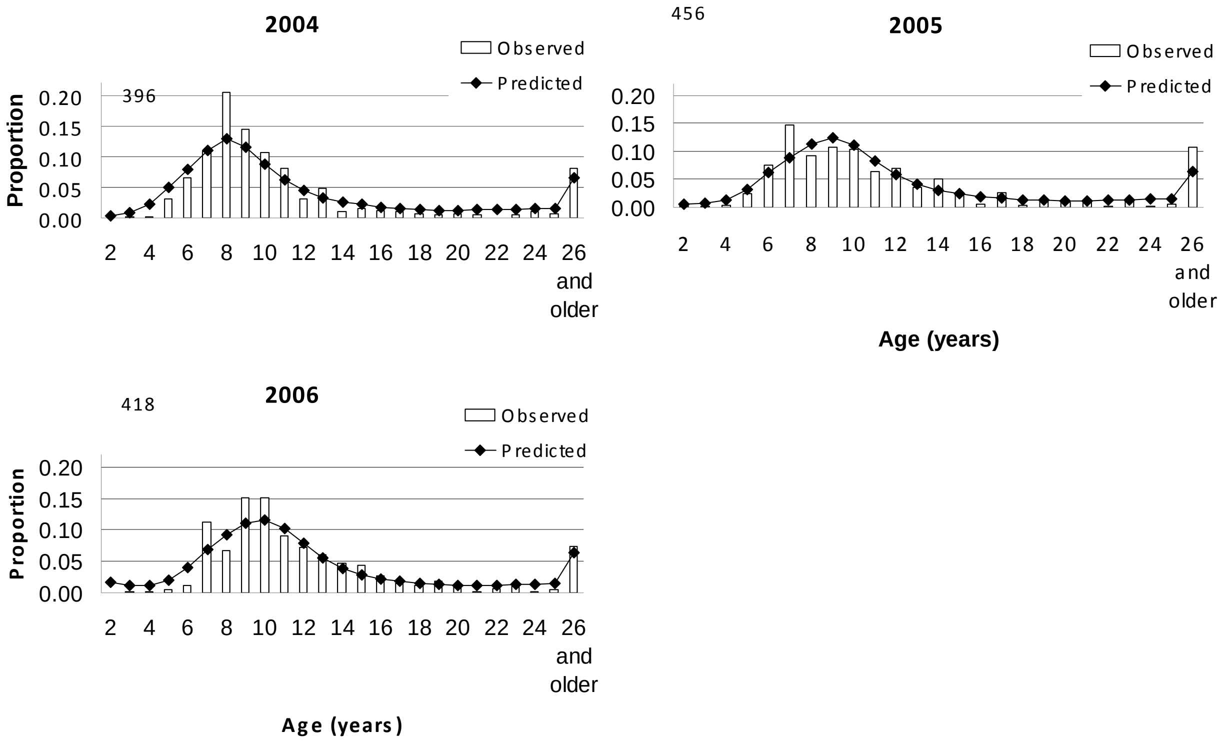


Figure 58.–Page 3 of 3.

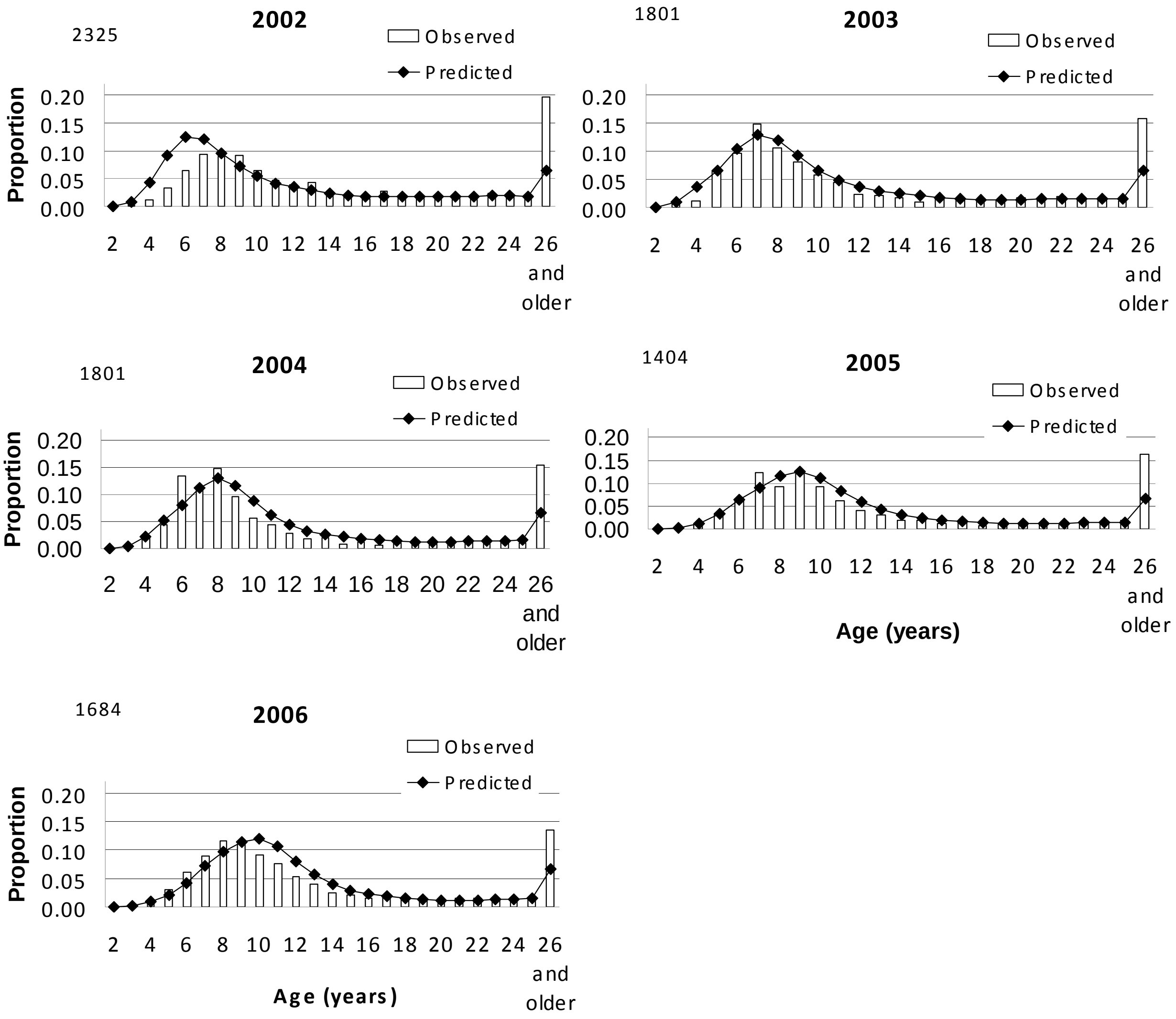


Figure 59.–Observed and predicted age composition during fishery, 2002–2006, under base model (*M* = 0.1, *q* for mark-recapture estimates fixed at 1). Numbers in the upper left corner of each figure denote sample size.

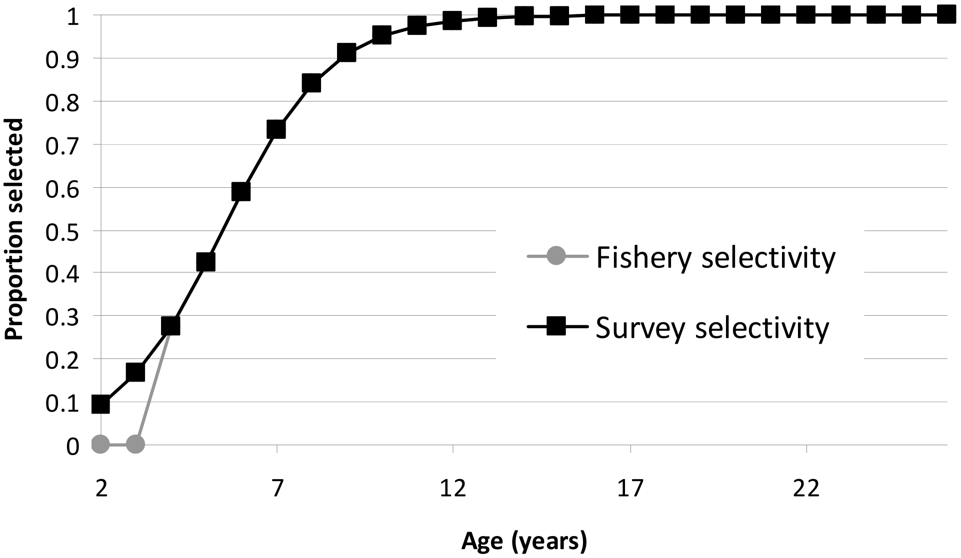


Figure 60.–Estimated fishery and longline survey selectivity of Chatham Strait sablefish under base model (*M* = 0.1, *q* for mark-recapture estimates fixed at 1).

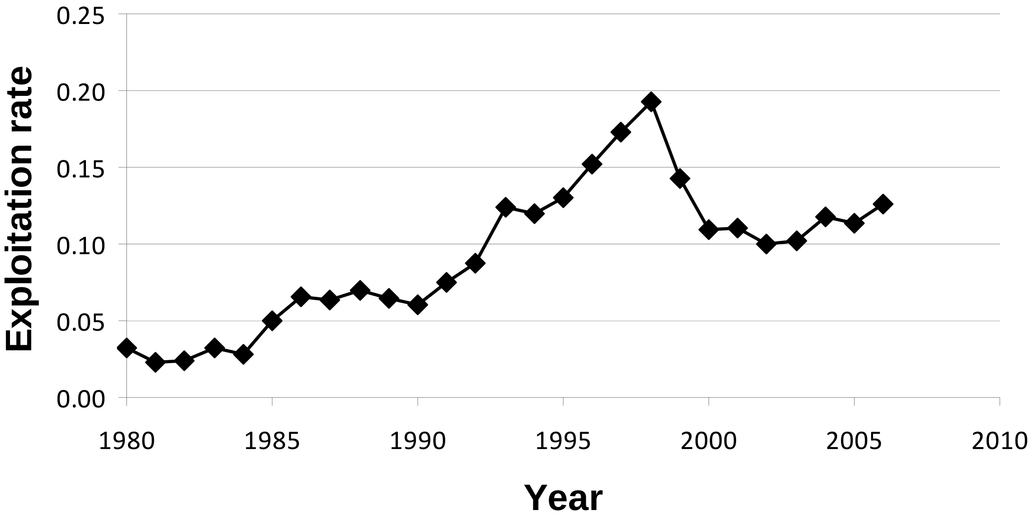


Figure 61.–Estimated exploitation rate (catch/exploitable biomass) of Chatham Strait sablefish, 1980–2006, under base model (*M* = 0.1, *q* for mark-recapture estimates fixed at 1).

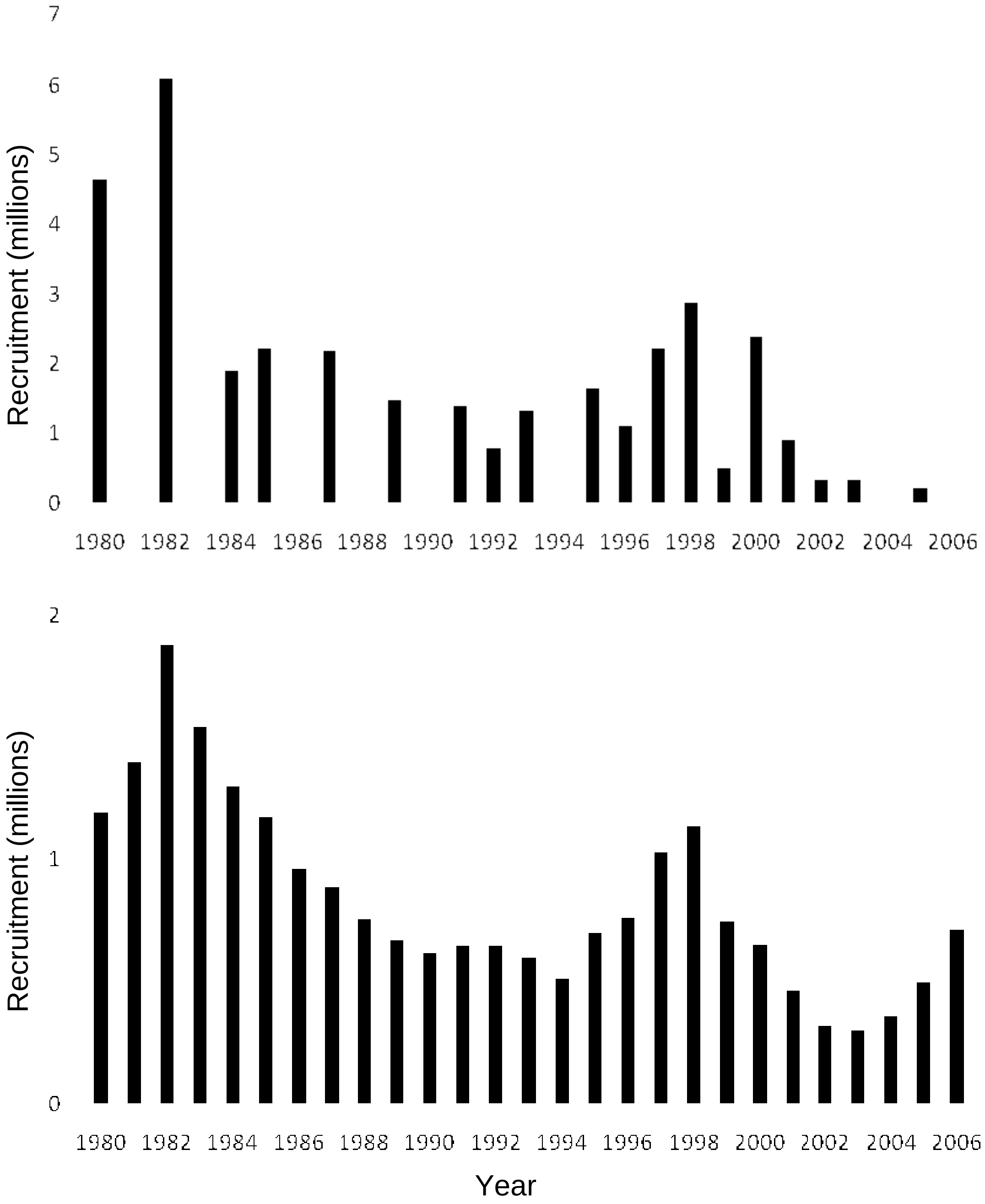


Figure 62.–Estimated recruitments, 1980–2006, for a model that did not constrain interannual variability in recruitment (top) and a model that strongly constrained interannual variability in R (bottom). Note that recruitment was estimated to be zero in a number of years ('81, '83, '86, '88, etc.) in the first case, which is most likely a result of allowing substantial ageing errors. Note similarity in overall trends.

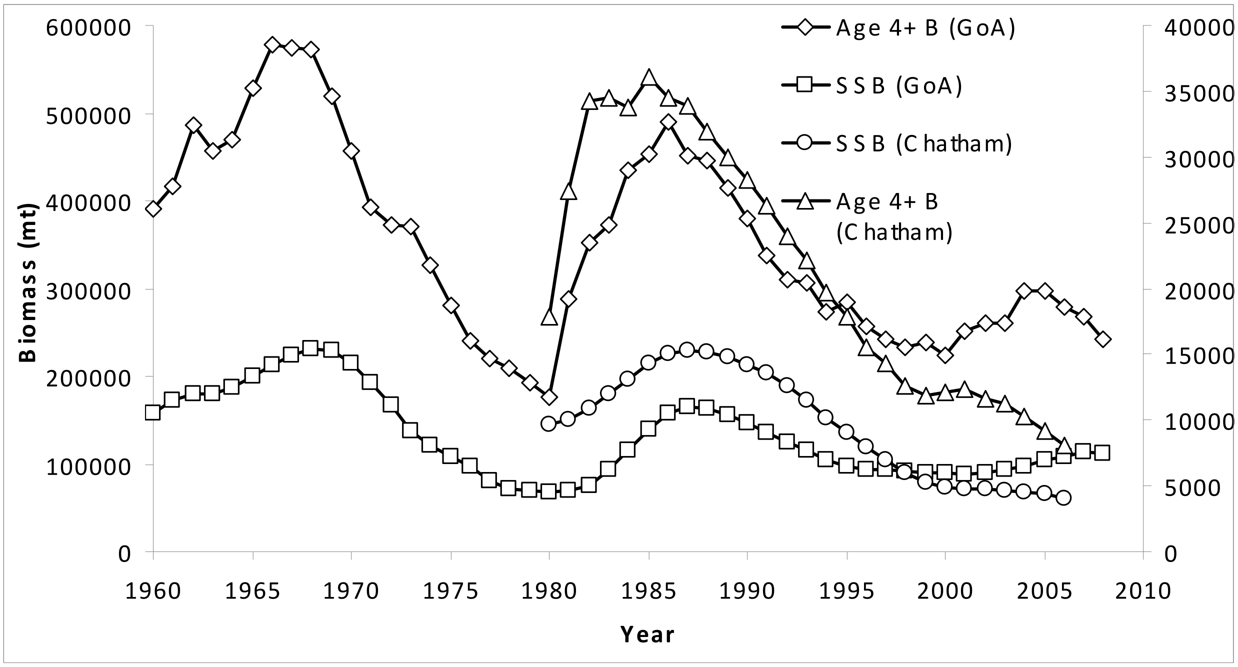


Figure 63.–Comparison of estimated biomass (Age 4+) and estimated spawning stock biomass (SSB) between the Gulf of Alaska/Bering Sea (as estimated by NMFS) and Chatham Strait (as estimated by base model).

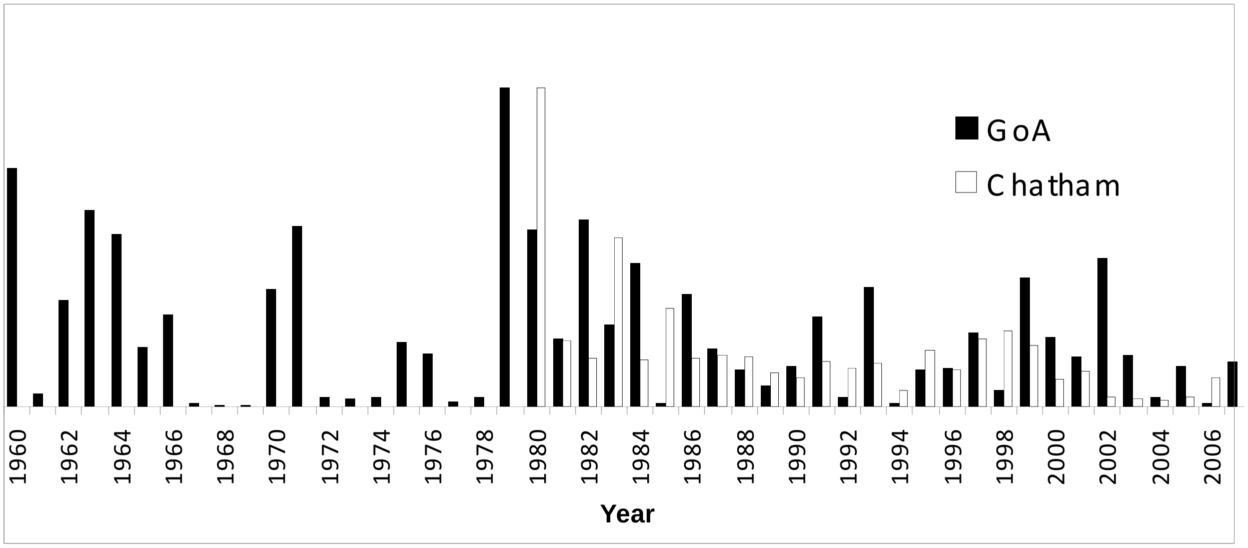


Figure 64.–Comparison of estimated relative recruitment time series between the Gulf of Alaska/Bering Sea (as estimated by NOAA) and Chatham Strait (as estimated by base model). Y-axis scale is arbitrary.

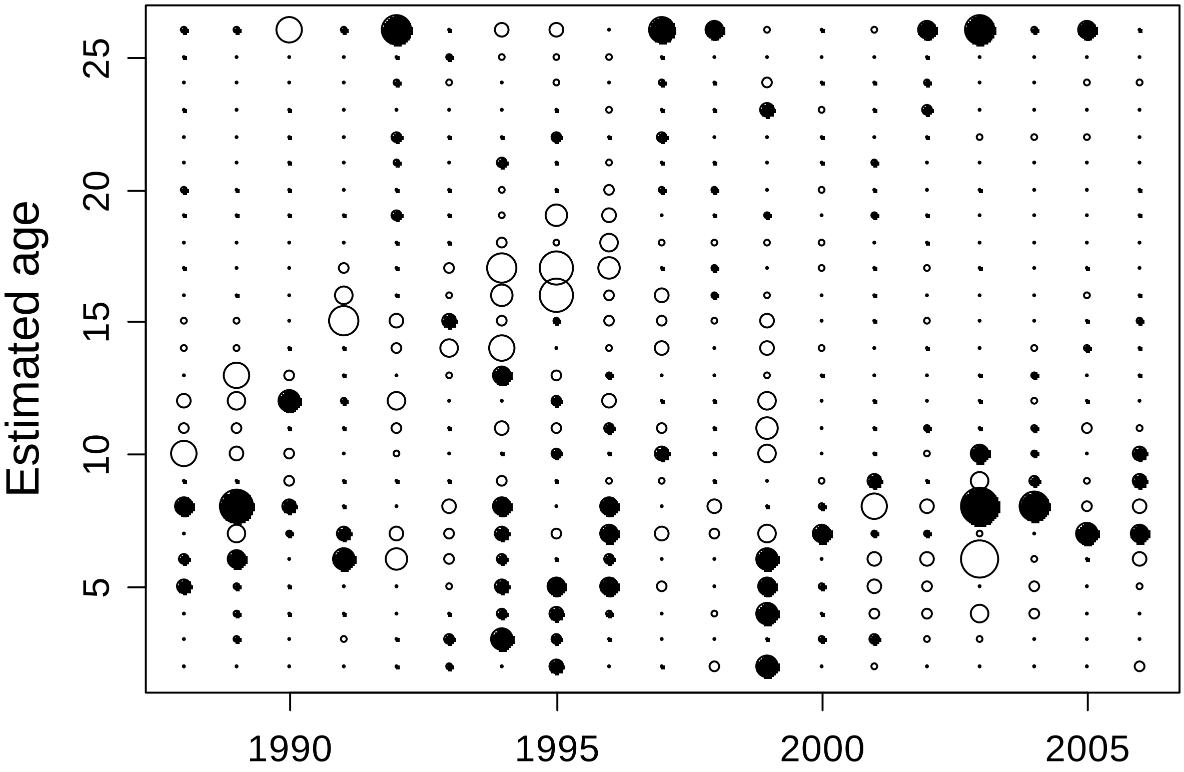


Figure 65.–Age composition residuals by year and age for longline survey. Filled circles denote positive residuals, open circles denote negative residuals.

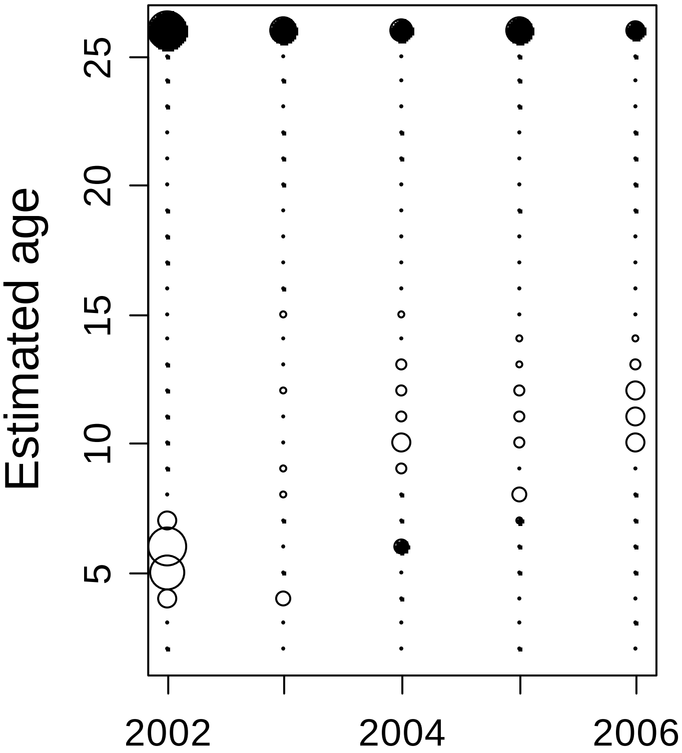


Figure 66.–Age composition residuals by year and age for fishery. Filled circles denote positive residuals, open circles denote negative residuals.

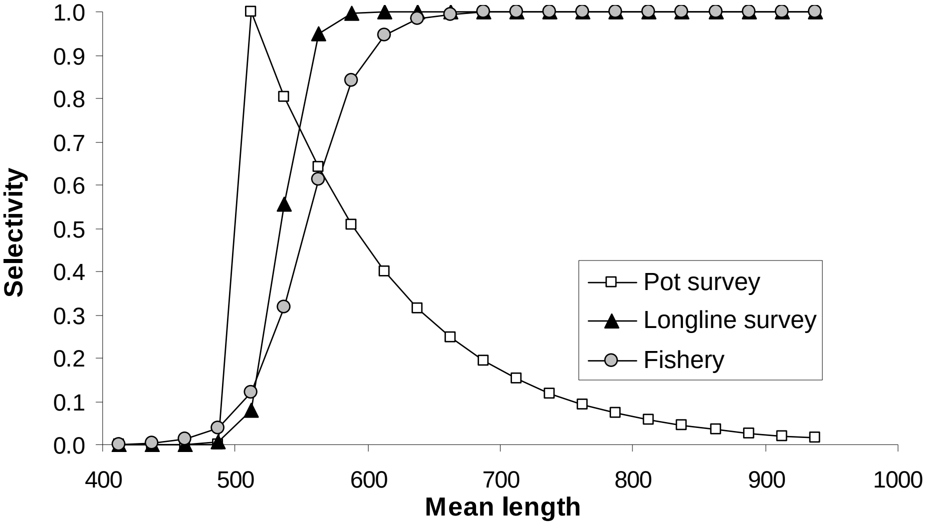
APPENDICES

Appendix A.1–Size selectivity simulations.

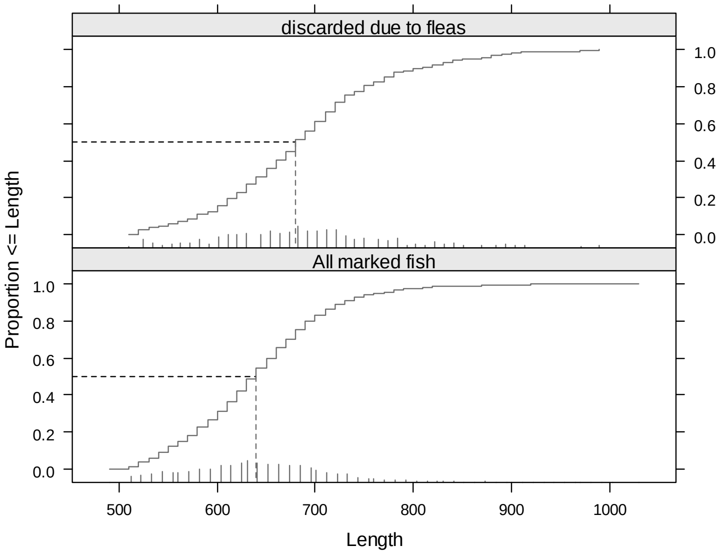
BACKGROUND

An important source of potential bias in the mark-recapture models may arise from fishery selectivity due to both gear selectivity (i.e., the inherent selectivity of the longline gear for certain sizes of fish) and additional fishery selectivity (i.e., “highgrading” by discarding small fish). There is strong evidence for considerable size selectivity in the fishery based on other fisheries and on the relative size composition of sablefish in the pot survey, longline survey, and longline fishery. For example, NOAA stock assessments for the Gulf of Alaska sablefish fishery estimate a pattern of selectivity-at-age that implies low selectivity at small sizes and full selectivity at larger sizes (logistic function). Size composition estimates in Chatham Strait from different gears employed in the past suggest that pot gear has a higher selectivity for smaller fish than longline gear and that the fishery selects against smaller fish relative to the longline survey, implying some highgrading or a different spatial distribution of the fishery (i.e., in areas with more larger fish, e.g., deeper stations on average). This pattern of size compositions was also observed in 2006 (Figure 19). It is unlikely that these differences are due to differences in spatial distribution because the depth distribution is similar among both surveys and the fishery (Figure 23), although the fishery extends into both shallower and deeper areas than the longline survey. In spite of the difference in apparent size selectivity between the fishery and the pot survey as evident in these figures, the size distribution of recovered (tagged) sablefish during the fishery is very similar to that of released fish from the pot survey (Figure 25), suggesting that all size classes that are marked by the pot survey are fully selected by the fishing gear.

The most parsimonious explanation of this observation seems to be that all size classes caught by pot gear and tagged are “vulnerable” to the fishery longline gear and that all tagged fish (which have external tags that are readily apparent when the longline is retrieved) are retained by fishermen, whereas untagged small fish tend to be discarded. I examined patterns in size selectivity by fitting a model to size composition data that estimates different selectivity curves for fish caught by the pot gear (and marked), for tagged fish caught in the fishery, and for the total retained catch. This model suggests full selectivity of tagged fish (i.e., near constant selectivity or probability of capture across all size classes that were tagged), a different selectivity pattern with a larger size at 50% selectivity for the total retained catch, and a strongly decreasing selectivity for sablefish by the pot gear (i.e., large sablefish have a much lower probability of being captured by the pot survey) (Appendix A.2). The model assumed asymptotic selectivity patterns for the longline gear and the fishery (following Hanselman et al. 2006), but allowed for dome-shaped selectivity in the pot survey because of the characteristics of the survey gear and the discarding of many large fish due to fleas (as evident in Appendix A.3). The size difference between marked and discarded fish in the pot survey was highly significant (Kolmogorov-Smirnov test, $D = 0.255$, $p < 0.001$) and strongly suggests that a substantial proportion of large fish were not marked, i.e., selected against by the pot survey. As noted above, the size composition of tagged fish retained in the fishery was the same size composition as that of fish tagged in the pot survey, suggests that the longline gear has full selectivity over the range of sizes of tagged fish. The size composition of the untagged fish retained in the fishery differs from that of the tagged fish, suggesting that only a portion of the untagged population that is caught by the gear is actually retained, most likely because of the discarding of small fish (highgrading). The effect of these selectivity patterns on mark-recapture estimates of population size was explored in simulations.



Appendix A.2–Estimated gear selectivity pattern for pot survey, longline survey, and fishery (retained catch after discarding small fish) based on size composition data.



Appendix A.3–Comparison of size composition of sablefish marked (clipped and/or tagged) during the pot survey and sablefish discarded due to fleas (not marked).

SIMULATION ANALYSES

I conducted a series of simulations to explore the effects of size selectivity, as well as variability in tag retention, on estimates of abundance. The simulations are based on breaking a hypothetical (but realistic) population down into 200 mm size classes, assuming that size selectivity in the longline fishery follows a logistic function, and computing the expected catch by size class (given a harvest rate of 10%) and the expected number of marked fish (clipped and tagged) recovered in the fishery. A variety of simple Petersen estimators were then computed under different scenarios to compare their performance. The simulations were done in an Excel spreadsheet and/or using R. The spreadsheet includes a number of build-in 'scenarios' that can be accessed through the 'Tools' menu and are described below. While only variations of the Petersen estimator for pooled data were examined, the results shown here are also applicable to other Petersen-type estimators such as the time-stratified binomial estimators examined in the 2006 assessment..

A comparison of the size composition of tagged fish recovered in the fishery and tagged fish released during the pot survey shows that their size composition is almost identical, which strongly suggests that all tagged fish (only fish greater than 560 mm were tagged in 2006) are fully selected for by the fishing gear, even at smaller sizes. However, differences between the size composition of all sablefish in the fishery and tagged sablefish recovered in the fishery suggest that, while all tagged fish appear to be retained, small untagged fish are not retained at the same rate (i.e., returned to the water - "highgrading"). Therefore the selectivity of tagged and untagged fish may differ substantially and they were modeled with separate selectivity curves in the simulations. A "gear selectivity" for the longline gear was applied to tagged fish, while a separate fishery selectivity (after "highgrading") was applied to all other fishes. Gear selectivity could possibly be estimated from the longline survey, which uses the same gear as the fishery but keeps fish of all sizes. NMFS estimates of selectivity-at-age of their longline survey (Hanselman et al. 2006) suggest that females are fully selected by the survey at age 4 (average size: 580 mm), therefore it seems reasonable that tagged fish over 560 mm may be fully selected for by the gear (if not by fishermen).

Based on observations of fishery operations, it seems unlikely that fish that are clipped only (i.e., have no tag) have a different selectivity than unclipped fish because fishermen shake small fish off the hook before they come aboard (hence it is unlikely that clipped fish can be identified and distinguished from unclipped fish). Therefore, for these simulations, I assumed that clipped and unclipped fish have the same selectivity (i.e., the same probability of capture, which also assumes that clipped fish are not selected *against*, e.g., by returning them to the water.).

Deterministic simulations were set up as follows:

1. I created a hypothetical "true" population based on the observed size distribution in the 2006 fishery, size selectivity estimated from NMFS data, and a harvest rate of 10%. Size selectivity was estimated by converting NMFS estimates of selectivity-at-age to selectivity-at-length using sex-specific LVB growth curves. The resulting estimates of selectivity-at-length differed between males and females and I simply averaged selectivity for each size class to obtain size selectivity. (Note: Selectivity-at-age from the Chatham Strait ASA model could be substituted for the NMFS estimates, although differences should be minor). The "hypothetical" population does not have any fish smaller than 500 mm because no smaller

fish are marked (even though some are caught in the pot survey). Therefore, all results only apply to the population of fish over 500 mm.

2. To simulate the pot survey sampling and generate samples of clipped fish and tagged fish (assuming fish greater than 560 mm are tagged), I assumed a harvest rate for the pot survey of 0.0025 (arbitrary, selected to obtain a sample size similar to the actual number of sablefish marked in 2006) and allowed the pot survey to have a dome-shaped selectivity (this was chosen somewhat arbitrarily as the 'base' scenario for most simulations based on evidence in the data that selectivity for larger fish is reduced, or is at least smaller than in the fishery). Selectivity was scaled to have a maximum of 1 (full selectivity for at least one size class).
3. Based on the assumed harvest rate and selectivity-at-size, "samples" of fish were generated and I assumed that all fish are clipped and fish greater than 560 mm (or any other cutoff value) are tagged.
4. Gear selectivity is assumed to follow a logistic function and is allowed to differ from fishery selectivity (which is assumed to be lower due to discarding of small fish). This accounts for the apparent higher selectivity of tagged fish (as evident in size composition data). For the 'base' scenario, I chose selectivity parameters that result in a selectivity very close to 1 for all fish larger than 560 mm (as suggested by size composition data).
5. Applying the assumed gear selectivity gives the size of the "vulnerable" population by size class. Summing across size classes results in the total vulnerable population for all size classes combined or for all fish greater than 560 mm.
6. The size selectivity in the fishery was assumed to follow a logistic function that, under the 'base' scenario, was as close as possible to the selectivity estimated from NMFS estimates of selectivity-at-age. Applying this selectivity to the true population gives estimates of the "exploited" population by size class, for all sizes combined, and for fish greater than 560 mm. Note that fishery selectivity is not constrained, but it would obviously not make sense to assume a higher selectivity at size in the fishery than for the longline gear.
7. Applying the 10% harvest rate and the assumed gear selectivity and fishery selectivity to the true population results in the expected number of fish caught by the longline gear and the retained catch, respectively.
8. The following expected numbers of clipped or tagged fish retained in the catch are calculated under different scenarios:
   a. The expected number of tagged fish retained in the fishery (only fish greater than 560 mm are tagged), assuming all tags caught by the gear are retained. Tag loss is accounted for using a tag loss/tag retention rate estimated from 2006 data (tag retention rate = number of tagged fish (356) divided by number of clipped fish (425)).
   b. The expected number of clipped fish (all sizes) retained in the fishery. This is a function of the selectivity of tagged fish (which equals gear selectivity) as well as of the selectivity of untagged fish. Assuming that all of the caught fish that had a tag were retained by fishermen, regardless of size, implies that clipped fish with a tag have the same selectivity as tagged fish, whereas fish with clips only (small fish and large fish that lost their tags) have the same selectivity as unmarked fish.
   c. The expected number of tagged fish retained in the fishery if all size classes of fish caught in the pot survey are tagged (i.e., all fish greater than 500 mm, as opposed to fish greater than 560 mm).

d. The expected number of clipped fish retained if NO tags were applied (i.e., retention of clipped fish is not affected by tags as in b.). Again, I assumed that clipped fish have the same selectivity, or probability of capture, as unmarked fish.

I constructed various Petersen estimators of abundance to examine their performance under the 'base' scenario and their sensitivity to uncertainty in the input Appendix A.4). All estimators were simple Petersen estimators of the exploited population (i.e., the fish that are retained by fishermen) because the samples (catch) corresponds to the exploited population. Estimators differed based on what kind of marks were used (clips or tags), what size classes were included (all fish greater than 500 mm or fish greater than 560 mm only), and whether they were corrected for selectivity or not. Two estimators (7 and 8) were based on the assumption that all clipped fish were tagged (rather than fish greater than 560 mm only, as in 2006) and four estimators (9–12) assumed that fish were clipped only, i.e., no tags were used (which changes their selectivity pattern). All estimators assumed that clipped (no tags) and unclipped fish have the same selectivity pattern (probability of capture), but the selectivity of tagged fish differs from that of untagged fish (because tags are easily visible and are assumed to be selected for, i.e., they have a higher probability of capture).

Appendix A.4–Input data and characteristics of various Petersen estimators of exploited sablefish abundance computed in simulations and results (percent deviation from "true" exploited population) under base scenario.

#	Marks	Size classes included	Corrected for selectivity?	Other	Over (+) or underestimation of true expl. pop./expl. pop. > 560 mm
1	tags	> 560 mm	No		-0.95%/0.01%
2	tags	> 560 mm	Yes		-0.96%/0.00%
3	clips	All	No		12.11%/13.19%
4	clips	All	Yes		0.00%/0.97%
5	clips	> 560 mm	No		0.95%/1.92%
6	clips	> 560 mm	Yes		-0.96%/0.00%
7	tags	all (all sizes tagged)	No		0.93%/1.90%
8	tags	all (all sizes tagged)	Yes		0.00%/0.97%
9	clips	All	No	No fish tagged	24.24%/25.44%
10	clips	All	Yes	No fish tagged	0.00%/0.97%
11	clips	> 560 mm	No	No fish tagged	11.98%/13.06%
12	clips	> 560 mm	Yes	No fish tagged	-0.96%/0.00%

The estimators are generally straightforward and are based on the simple Petersen estimate: $N = M * n/m$, where N is the estimated population size (either all fish or large fish over 560 mm), M is the number of marks (clips and/or tags) on which the estimator is based and which are "available" to the fishery, n is the sample size (equal to the number of fish caught, assuming all fish are examined for marks), and m is the number of marks in the catch. For estimators based on tags (1, 2, 7, and 8), the "available" number of tags depends on the tag retention rate and is computed as the product of tag retention, r, and the original number of tags, M_t: $M_t' = r * M_t$. If gear selectivity is less than 1 at any size, the number of available tags is further discounted by multiplying M_t' by an appropriate correction factor. The correction factor is a function of selectivity at size and the number of tags in the marked population at a given size. If $s_{t,i}$ is the selectivity for tagged fish in size group i (equal to gear selectivity prior to discarding small fish) and $M'_{t,i}$ is the number of remaining tagged fish in size group i (after tag loss), the correction factor f_t is the sum of their products over all k size groups, divided by the total number of remaining tags:

$$f_t = \sum_{i=1}^{k} s_{t,i} \cdot M'_{t,i} \Big/ \sum_{i=1}^{k} M'_{t,i}$$

The correction factor reduces the number of "available" tags to $f_t * r * M_t$ and the estimated population size corrected for size selectivity of the longline gear (estimators 2 and 8) is therefore computed based on the number of recovered tags (m_t) as follows:

$$N = (f_t * r * M_t) * n/m_t$$

Estimators 4 and 6 are based on the number of clipped fish recovered (m_c), for which the number of "available" clipped fish depends on both the number of available tagged fish ($f_c * r * M_t$), all of which also have clips, and on the number of available fish in the population that are clipped but have no tags. The latter depends on tag retention rate and on fishery selectivity (after discarding small fish). The correction factor for clipped fish, f_c, is computed as above by replacing the selectivity of tagged fish in size group i, $s_{t,i}$, with the selectivity of clipped fish (equals fishery selectivity) in size group i, $s_{c,i}$, and replacing the number of tags in a size group with the number of clips in a given size group ($M'_{c,i}$):

$$f_c = \sum_{i=1}^{k} s_{c,i} \cdot M'_{c,i} \Big/ \sum_{i=1}^{k} M'_{c,i}$$

The number of fish in the population with clips only is determined by the number of clipped fish that are not tagged in the first place (fish less than 560 mm) plus the number of tagged fish that lose their tags, or:

$$(M_c - M_t) + (1-r) * M_t = M_c - r * M_t$$

The total number of clipped fish that are "available" to the fishery is then determined by the number of tagged fish available plus the number of clipped fish available:

$$M_c = f_t * r * M_t + f_c\,(M_c - r * M_t)$$

and population size is estimated based on the number of available marks and the number of recovered marks as $N = M_c * n/m_c$

To test the simulation model and compare estimators I first assumed that fishery selectivity and the tag loss/retention rate are known without error (perfect information) and computed estimates under the 'base' scenario (which reflects "best guess" selectivity scenarios for the pot survey, longline gear, and the fishery). Secondly, I explored the effects of mis-specification of gear and fishery selectivity and of tag retention on the results. Under the base scenario and with perfect information, the following results were obtained:

- The correction for fishery selectivity actually works (confirming that formulas seem to be correct). In all cases where it was applied, the adjustment results in a correct estimate of the size of the exploited population (given that gear selectivity and/or fishery selectivity and tag retention rate are known without error.)
- Uncorrected estimates based on tag returns only slightly overestimated the true size of the exploited population (0.01% to 0.92%) under the base scenario.

- Estimator 1 uses the number of tags (fish greater than 560 mm) returned from the fishery (*m*), the number of tagged fish in the population discounted by tag loss ($M' = r * M$, where M is the number of tagged fish released and r is the tag retention rate estimate from 2006 data), and the "sample size" n (the number of fish caught in the fishery). It results in a very accurate estimate of the size of the exploited population of fish over 560 mm under the assumptions of the base model (most notably that gear selectivity for fish over 560 mm is close to or equal to 1) and with perfect information (i.e., known tag retention rate). At first glance it seemed surprising that this actually provides an estimate of exploited population size, given that the tagged fish had a higher probability of being retained than untagged fish in the model. However, because the estimator uses a "sample size" n that is based on the number of fish retained rather than the number of fish caught prior to discarding of small fish, it provides an estimate of the exploited population rather than the "vulnerable" population as can be seen in the following ratios. The Petersen estimator is based on the fact that the ratio of marks in the population (M/N) is the same as the ratio of marked recaptures in the sample (m/n) IF the capture probabilities are the same for marked and unmarked fish (which they are not in the case of tagged sablefish). Because the "capture" probability (rather, the "retention" probability in the fishery) is higher for tagged fish than for untagged fish, it may seem that the "inflated" number of marks would result in an underestimate of population size. However, as long as the capture probability by the gear (i.e., gear selectivity prior to discarding) is the same for tagged and untagged fish, we still get an accurate estimate of the exploited population size. The difference in capture probability (i.e., selectivity) leads to a sample size (n_{expl} = retained catch) that is lower then the number of fish actually caught by, or vulnerable to the gear (n_{vuln}). Because of the fact that the ratio n_{expl}/n_{vuln} is the same as the ratio of the exploited population to the vulnerable population (N_{expl}/N_{vuln}), the Petersen estimator based on n_{expl} results in an estimate of exploited abundance:

Basis for Petersen estimator: $$\frac{m}{n_{vuln}} = \frac{M}{N_{vuln}}$$

rearranging: $$\frac{m}{M} = \frac{n_{vuln}}{N_{vuln}}$$

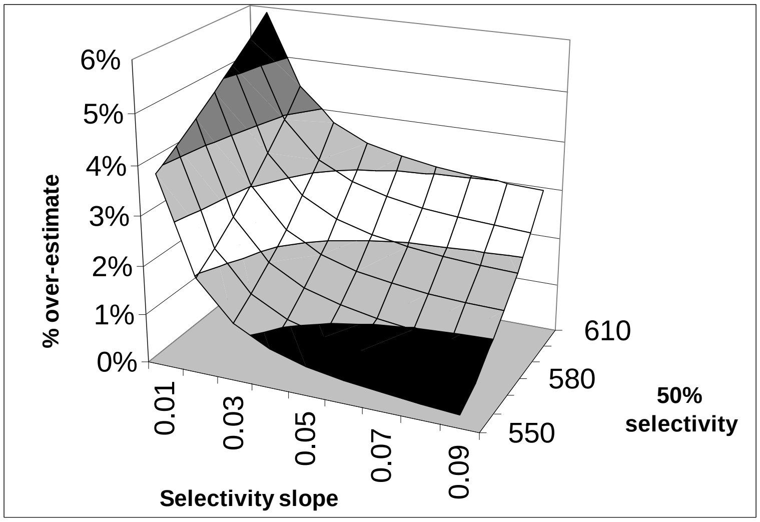
substituting: $$\frac{m}{M} = \frac{n_{expl}}{N_{expl}}$$

hence the usual Petersen estimator yields:

$$N_{expl} = \frac{n_{expl} \cdot M}{m}$$

- Estimator 7, which is very similar but assumes that all clipped fish were tagged, results in a somewhat higher (but still very low) overestimate of about 1% of the exploited population. The overestimate (compared to estimator 1) results from the built-in assumption that gear selectivity for fish less than 560 mm is less than 1 under the base scenario. Of course, we know little about actual gear selectivity at smaller sizes and hence the amount of overestimation is very uncertain. Under the extreme case that gear selectivity is zero for fish less than 560 mm and selectivity is 1 for fish 560 mm and larger, this estimator overestimates the true exploited population size by 11% with all other parameters as in the 'base' scenario.

- The uncorrected estimator based on the number of clipped fish that were greater than 560 mm length (estimator 5) overestimated exploited population size by only 1.92% under the assumed base scenario, which assumed a size at 50% selectivity of 586 mm (slope = 0.0466). Variability in the slope and size at 50% selectivity resulted in the following amounts of overestimation (assuming tag retention rate is known without error and tagged fish are fully selected by gear). The amount by which exploited population size is overestimated increases approximately linearly with the 50% selectivity parameter and rises rapidly as the slope of the selectivity curve decreases (i.e., gets less steep). However, over most of the "likely" range of values, the exploited population size (fish greater than 560 mm) is overestimated by less than 2–3% (Appendix A.5).



Appendix A.5–Sensitivity of estimator 5 to variability in fishery selectivity. Exploitable population size based on the uncorrected estimator is over-estimated by 0–6% over the range of selectivities examined.

Other uncorrected estimators (based on clipped fish) resulted in considerable over-estimates of exploited population size under the base scenario as a direct consequence of size selectivity (12–24%, Appendix A.4), while they either over- or (more typically) underestimated the vulnerable population size.

- The uncorrected estimators accurately estimate exploited population size only if no size selectivity occurs in the longline fishery (i.e., fish of all sizes, as well as tagged and untagged fish, have the same probability of capture). In that case, the exploited population is identical to the vulnerable population and the exploited population size is accurately estimated **regardless of the selectivity pattern of the pot survey**. With increasing levels of size selectivity, the bias in the uncorrected estimators increases. For example, under the base scenario for pot selectivity and gear selectivity, the amount of overestimation increases the more selective the fishery becomes.

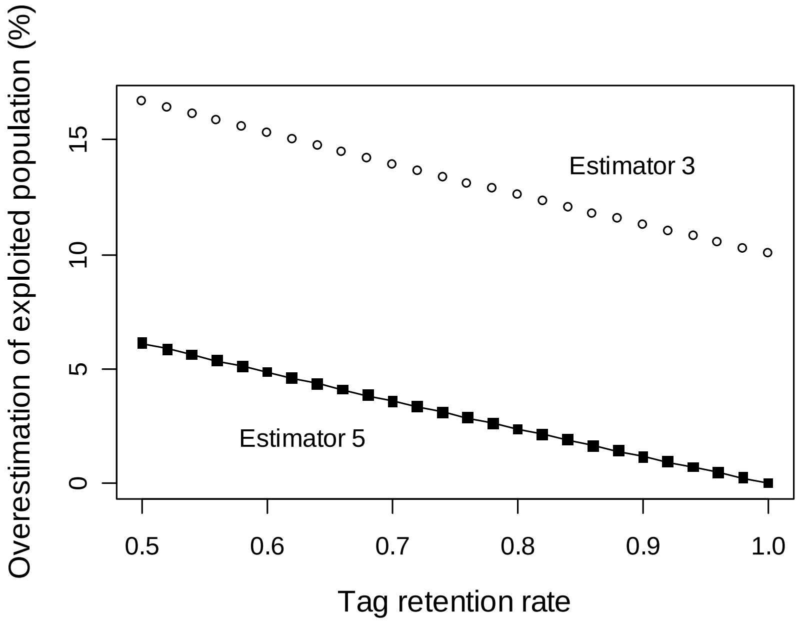
In general, the simulations show that the Petersen estimators based on clips or tags lead to an overestimate of exploited population size, if the estimators are not corrected for size selectivity. The uncorrected estimators underestimated the vulnerable population in most cases although the estimator based on clips for the case where NO fish are tagged even overestimated the vulnerable population size. The size of the total vulnerable population (fish greater than 500 mm) is difficult to estimate because it depends on the difference in selectivity between the gear *per se* and the fishery (i.e., on the number of small fish discarded). Because selectivity at small sizes is poorly known and difficult to estimate based on the small number of fish in those size classes, estimators of population size above 560 mm were generally much less sensitive to overestimation.

RESULTS

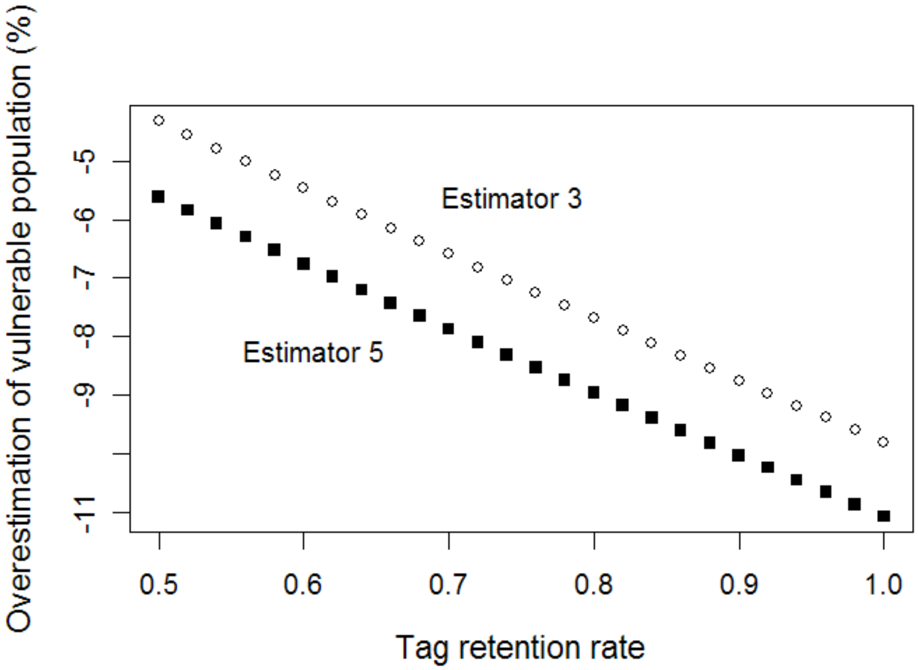
This section summarizes results from deterministic simulations under various scenarios:

1. Effect of tag retention with "perfect information" (i.e., true tag retention rate and selectivity schedules are known)

Only estimators 3 and 5 (based on all/large clipped fish, respectively) are sensitive to variability in tag retention rate, because the tag retention rate affects the number of clipped fish retained in the fishery. For example, if tag loss is high, fewer tagged fish are encountered and kept, and therefore the number of clipped fish retained will be lower (because of the discarding of small fish with clips only, whereas small tagged fish are assumed to be kept). Other estimators are not affected by or correct for the effects of tag retention (if known). The effects of tag retention on estimators 3 and 5 as estimators of exploited population size Appendix A.6) and as estimators of vulnerable population size (Appendix A.7) vary linearly with tag retention rate because the number of “available” marks in the population is a linear function of tag retention rate. Both estimators 3 and 5 lead to a much more severe over-estimation of exploited population size at lower tag retention rates, but to a less severe under-estimation of vulnerable population size at lower tag retention rates. Under the ‘base’ scenario and at the 2006 estimated tag retention rate (if known without error), estimator 3 (based on all clips) would <u>overestimate</u> exploited population size by 12% and <u>underestimate</u> vulnerable population size by 8%, while estimator 5 (based on clipped fish greater than 560 mm) would <u>overestimate</u> exploited population size (fish greater than 560 mm) by < 2% and <u>underestimate</u> vulnerable population size by 9%. **Therefore, of those estimators that are sensitive to tag retention rate estimator 5, if considered an estimator of <u>exploited</u> population size, appears to be least sensitive to variability in tag retention.** However, if tag retention (and gear/fishery selectivity) is known accurately, all other estimators provide unbiased estimates of exploited population size.



Appendix A.6–Effect of variability in tag retention rate on estimators 3 and 5 as estimators of exploited population size. Exploited population size is overestimated at all levels of tag retention, except at 100% retention.



Appendix A.7–Effect of variability in tag retention rate on estimators 3 and 5 as estimators of vulnerable population size. Vulnerable population size is underestimated at all levels of tag retention.

2. Effect of pot survey selectivity on estimates, given "perfect information"

The effects of pot survey selectivity were examined by varying the selectivity pattern of the pot survey and monitoring the effect on the various estimators. Pot survey selectivity was "bracketed" by two extreme scenarios (a, b) and one intermediate scenario (c). Scenarios were:

a. Assume that selectivity is much higher at younger ages and decreases with age, i.e., the proportion of small fish clipped is much larger than the proportion of larger fish that are clipped (*dome*)
b. Assume that selectivity is constant at all sizes (*sel = 1*)
c. Assume a logistic selectivity function similar to the fishery but shifted towards younger ages (*logistic*)
d. Assume a 'standard' scenario ('*std'*, used for all other simulations) with intermediate selectivity pattern (peak at 535 mm, decreasing selectivity with size to ~ 0.4 at 1000 mm).

Appendix A.8–The relative deviations (in percent) from the "true" exploited abundance under each scenario with brief descriptions of each estimator.

Estimator	sel = 1	Logistic	Dome	Std	Marks	Sizes	Corrected?	Other
1	0.00%	0.00%	0.00%	0.01%	tags	> 560 mm	No	
2	0.00%	0.00%	0.00%	0.00%	tags	> 560 mm	Yes	
3	11.50%	9.55%	16.13%	12.09%	clips	all	No	
4	0.00%	0.00%	0.00%	0.00%	clips	all	Yes	
5	1.83%	1.80%	2.40%	1.92%	clips	> 560 mm	No	
6	0.00%	0.00%	0.00%	0.00%	clips	> 560 mm	Yes	
7	0.10%	0.06%	0.14%	0.92%	tags	all	No	
8	-0.04%	-0.08%	0.00%	0.00%	tags	all	Yes	
9	23.03%	20.68%	32.30%	24.22%	clips	all	No	No fish are tagged
10	0.00%	0.00%	0.00%	0.00%	clips	all	Yes	No fish are tagged
11	12.46%	12.23%	16.85%	13.06%	clips	> 560 mm	No	No fish are tagged
12	0.00%	0.00%	0.00%	0.00%	clips	> 560 mm	Yes	No fish are tagged

Results suggest that most estimators, if considered to estimate exploited population size, are not very sensitive to fairly extreme differences in pot survey selectivity (Appendix A.8). Generally, estimators that were corrected for size selectivity were not affected by variations in survey selectivity. Similarly, estimators based on tagged fish were largely unaffected. Stronger effects were evident for the uncorrected estimators using clipped fish. While there were relatively small differences between the logistic and constant selectivity scenarios, in the presence of extreme dome-shaped selectivity these estimators led to a more pronounced overestimation of exploited biomass.

All estimators except those (uncorrected) estimators that assume that NO fish are tagged (hence different selectivity of tagged fish does not affect number of marks recovered) result in considerable underestimates of vulnerable population size (Appendix A.9). However, estimators 9 and 11 result in nearly unbiased estimates of vulnerable population size under a wide range of scenarios for pot survey selectivity. Only if pot survey selectivity is assumed to be 1 for all size classes (greater than 500 mm), these estimators lead to a considerable overestimate of population size (Appendix A.9, + 8.5% and + 3.9%, respectively).

Appendix A.9–The relative deviations (in percent) from the "true" vulnerable abundance under each scenario with brief descriptions of each estimator.

Estimator	sel = 1	Logistic	Dome	Std	Marks	Sizes	Corrected?	Other
1	-11.1%	-11.1%	-11.1%	-11.1%	tags	> 560 mm	No	
2	-11.1%	-11.1%	-11.1%	-11.1%	tags	> 560 mm	Yes	
3	-4.8%	-8.6%	-10.2%	-8.1%	clips	all	No	
4	-18.0%	-18.0%	-18.0%	-18.0%	clips	all	Yes	
5	-8.9%	-9.4%	-9.5%	-9.4%	clips	> 560 mm	No	
6	-11.1%	-11.1%	-11.1%	-11.1%	clips	> 560 mm	Yes	
7	-17.0%	-17.3%	-17.5%	-17.3%	tags	all	No	
8	-18.0%	-18.0%	-18.0%	-18.0%	tags	all	Yes	
9	8.5%	0.9%	-1.1%	1.9%	clips	all	No	No fish are tagged
10	-18.0%	-18.0%	-18.0%	-18.0%	clips	all	Yes	No fish are tagged
11	3.9%	0.0%	-0.2%	0.5%	clips	> 560 mm	No	No fish are tagged
12	-11.1%	-11.1%	-11.1%	-11.1%	clips	> 560 mm	Yes	No fish are tagged

3. Effect of longline (gear) selectivity on estimates under "perfect information"

The effect of variability in the selectivity of the longline gear (prior to discarding) on estimates was examined using three scenarios that are likely to bracket the true pattern of selectivity:

a. Assume that gear selectivity is similar to the selectivity estimated by NMFS for the domestic longline survey.
b. Assume that gear selectivity is constant for fish of all sizes greater than 500 mm (*sel* = 1.)
c. Assume intermediate scenario ('base') with a logistic selectivity function that has selectivity close to 1 for fish greater than 560 mm (as suggested by size composition data). This scenario was used in all other simulations.

Appendix A.10–The relative deviations (in percent) from the "true" exploited abundance under each gear selectivity scenario with brief descriptions of each estimator.

Estimator	sel = 1	Base	NMFS	Marks	Sizes	Corrected?	Other
1	0.00%	0.01%	7.11%	tags	> 560 mm	No	
2	0.00%	0.00%	0.00%	tags	> 560 mm	Yes	
3	12.08%	12.09%	18.75%	clips	all	No	
4	0.00%	0.00%	0.00%	clips	all	Yes	
5	1.91%	1.92%	8.04%	clips	> 560 mm	No	
6	0.00%	0.00%	0.00%	clips	> 560 mm	Yes	
7	0.01%	0.92%	15.81%	tags	all	No	all clipped fish are tagged
8	0.00%	0.00%	0.00%	tags	all	Yes	all clipped fish are tagged
9	24.22%	24.22%	24.22%	clips	all	No	No fish are tagged
10	0.00%	0.00%	0.00%	clips	all	Yes	No fish are tagged
11	13.06%	13.06%	13.06%	Clips	> 560 mm	No	No fish are tagged
12	0.00%	0.00%	0.00%	Clips	> 560 mm	Yes	No fish are tagged

Gear selectivity affects all uncorrected estimators of exploited abundance if fish are tagged (estimators are unaffected if no fish are tagged: see 9–12 in Appendix A.10). Under the '*base*' scenario we assume that selectivity is close to 1 for all size classes above 560 mm, hence there is little differences between scenarios '*base*' and '*sel* = 1'. However, if gear selectivity is similar to fishery selectivity (i.e., no or little discarding under the NMFS selectivity scenario), uncorrected estimators based on tags (1 and 7) may severely overestimate exploited population size. Similarly, uncorrected estimators based on clips (3 and 5) lead to a more pronounced overestimation compared to the other scenarios. Estimates of population size over 560 mm (1 and 5) are much less susceptible to lower gear selectivity because even under the '*NMFS*' scenario most fish over 560 mm are fully recruited to the gear.

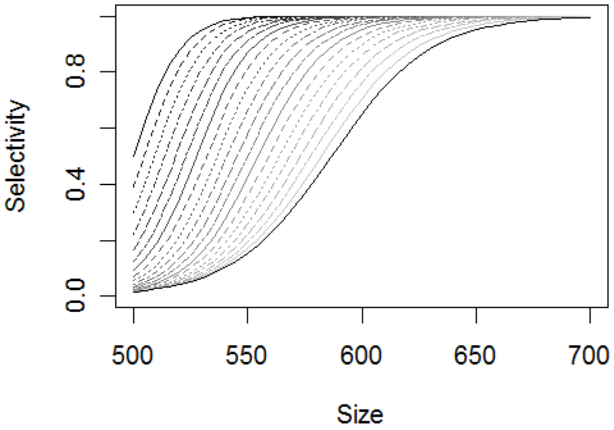
As estimators of vulnerable population size, all estimators except 9 and 11 lead to considerable underestimates of abundance under the base scenario and under the selectivity = 1 scenario (i.e., all size classes fully recruited to the gear) (Appendix A.11). In contrast, estimators 9 and 11 are relatively unbiased under these scenarios, but lead to considerable overestimates of abundance (16% and 7.4%, respectively) if longline selectivity is assumed to be the same as that estimated by NMFS for the domestic longline survey (which was considered an "extreme" scenario here). I'm not exactly sure why 9 and 11 overestimate vulnerable population size, but they are really estimators of total population and provide an unbiased estimate of the total (hypothetical) population size if fish are fully recruited to the pot survey gear (i.e., if the number of clips in each size class is proportional to the true population). The amount of overestimation seems to depend in complex ways on the relative patterns of selectivity among the different gears.

Appendix A.11–The relative deviations (in percent) from the "true" vulnerable abundance under each gear selectivity scenario with brief descriptions of each estimator.

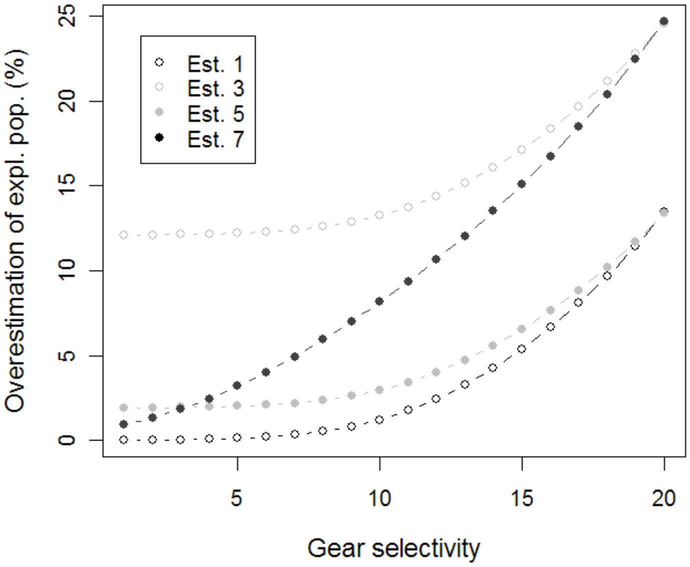
Estimator	sel = 1	Base	NMFS	Marks	Sizes	Corrected?	Other
1	-11.1%	-11.1%	1.7%	Tags	> 560 mm	No	
2	-11.1%	-11.1%	-5.0%	Tags	> 560 mm	Yes	
3	-8.9%	-8.1%	11.1%	Clips	all	No	
4	-18.7%	-18.0%	-6.5%	Clips	all	Yes	
5	-9.4%	-9.4%	2.6%	Clips	> 560 mm	No	
6	-11.1%	-11.1%	-5.0%	Clips	> 560 mm	Yes	
7	-18.7%	-17.3%	8.3%	Tags	all	No	all clipped fish are tagged
8	-18.7%	-18.0%	-6.5%	Tags	all	Yes	all clipped fish are tagged
9	1.0%	1.9%	16.2%	Clips	all	No	No fish are tagged
10	-18.7%	-18.0%	-6.5%	Clips	all	Yes	No fish are tagged
11	0.5%	0.5%	7.4%	Clips	> 560 mm	No	No fish are tagged
12	-11.1%	-11.1%	-5.0%	Clips	> 560 mm	Yes	No fish are tagged

I examined sensitivity of estimators 1, 3, 5, and 7 to variability in gear selectivity further by computing estimates over a range of gear selectivity that varied smoothly between the gear selectivity and fishery selectivity assumed under the base scenario (Appendix A.12).

Results suggest that, among the uncorrected estimators, estimator 3 (based on all clips) is clearly most strongly affected by gear selectivity and overestimates exploited population size by 12–25% over the range of values examined, whereas estimators 1 and 5 (based on larger fish only) are least sensitive to gear selectivity (Appendix A.13).



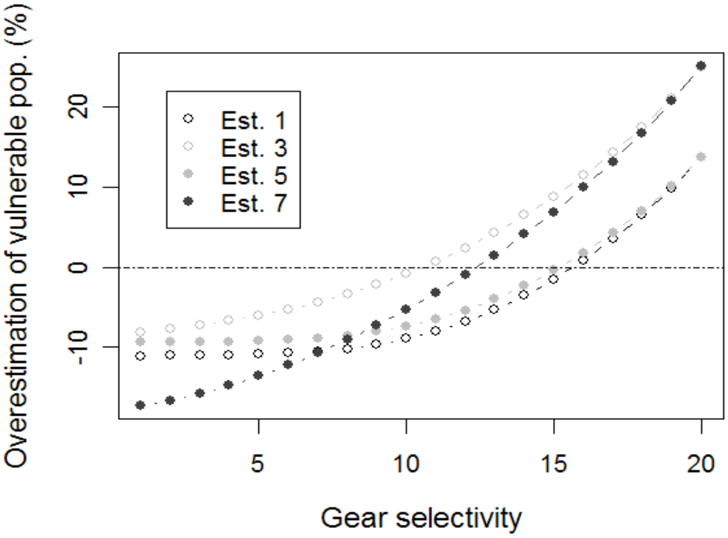
Appendix A.12–Range of gear selectivities used to examine sensitivity of estimator to variability in gear selectivity.



Appendix A.13–Effect of gear selectivity on uncorrected estimators (of exploited population size) that are sensitive to variations in gear selectivity.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12, from left to right.

Relative to vulnerable population size, the estimators all behave very similarly and are much more sensitive to variability in gear selectivity with much more extreme under and over-estimation of vulnerable population size (Appendix A.14)



Appendix A.14–Effect of gear selectivity on uncorrected estimators (of vulnerable population size) that are sensitive to variations in gear selectivity.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12, from left to right.

4. Effect of fishery selectivity on estimates under "perfect information"

The effect of variability in fishery selectivity on estimator 5 (uncorrected estimator based on the number of clipped fish greater than 560 mm) was already summarized above in Appendix A.5. Here, I first examined the sensitivity of the estimators under various scenarios that span the range of reasonable scenarios:

a. assume that gear selectivity is equal to the "standard" gear selectivity under the base scenario (i.e., no discarding of small fish) (scenario 'none').
b. assume an intermediate 'base' scenario with a logistic selectivity function that was based on NMFS estimates of selectivity in the IFQ longline fishery (50% selectivity at 586 mm). This scenario was used in all other simulations.
c. assume a more extreme scenario with 50% selectivity at 620 mm and a slope of 0.02
d. Note: right-most curve in Appendix A.17. This was considered an extreme scenario reflecting very pronounced discarding of small sablefish.

Exploited population size is overestimated by all uncorrected estimators under all three scenarios Appendix A.15. With no fishery selectivity ('none' beyond selectivity of gear, i.e., no discarding of small fish) none of the estimators over-estimated population size by more than 1% (assuming that true selectivities and tag retention rate are known). However, over-estimation increases very strongly with more pronounced fishery selectivity (relative to gear selectivity, i.e., under pronounced “highgrading”). Estimators 9 and 11 (as well as 3) were most strongly affected by highgrading.

Appendix A.15–The relative deviations (in percent) from the “true” exploited abundance under each fishery selectivity scenario with brief descriptions of each estimator.

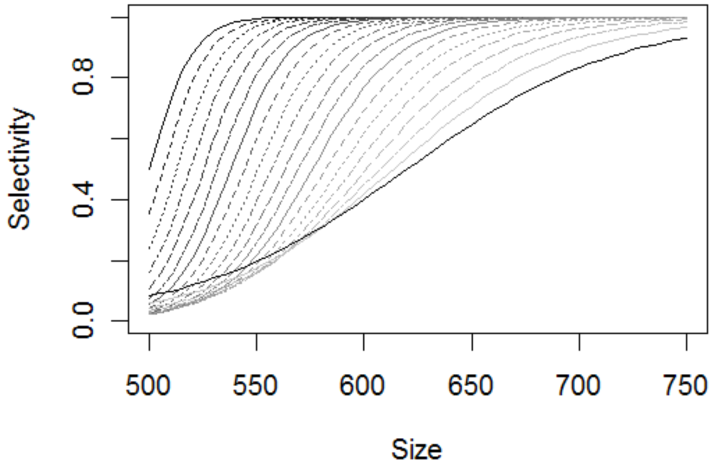
Estimator	None	Base	Extreme	Marks	Sizes	Corrected?	Other
1	0.01%	0.01%	0.01%	Tags	> 560 mm	No	
2	0.00%	0.00%	0.00%	Tags	> 560 mm	Yes	
3	0.93%	12.11%	14.59%	Clips	all	No	
4	0.00%	0.00%	0.00%	Clips	all	Yes	
5	0.01%	1.92%	4.97%	Clips	> 560 mm	No	
6	0.00%	0.00%	0.00%	Clips	> 560 mm	Yes	
7	0.93%	0.93%	0.93%	Tags	All	No	all clipped fish are tagged
8	0.00%	0.00%	0.00%	Tags	All	Yes	all clipped fish are tagged
9	0.93%	24.24%	53.14%	Clips	All	No	No fish are tagged
10	0.00%	0.00%	0.00%	Clips	All	Yes	No fish are tagged
11	0.01%	13.07%	41.10%	Clips	> 560 mm	No	No fish are tagged
12	0.00%	0.00%	0.00%	Clips	> 560 mm	Yes	No fish are tagged

In contrast, vulnerable population size is underestimated by all uncorrected estimators if fishery selectivity is present (scenarios ‘base’ and ‘extreme’ in Appendix A.16). With no fishery selectivity (‘none’ beyond selectivity of gear, i.e., no discarding of small fish), estimates of exploitable and vulnerable population size are identical, hence column ‘none’ in Appendix A.15 and Appendix A.16 are identical. The magnitude of under-estimation of vulnerable population size is generally much higher than the amount of over-estimation of exploited population size for most estimators. However, estimators 9 and 11 provide relatively accurate estimates of vulnerable population size and are not very sensitive to highgrading (Appendix A.16). They result in slight overestimates of vulnerable abundance under all three scenarios.

Appendix A.16–The relative deviations (in percent) from the “true” vulnerable abundance under each fishery selectivity scenario with brief descriptions of each estimator.

Estimator	None	Base	Extreme	Marks	Sizes	Corrected?	Other
1	0.01%	-11.1%	-28.2%	Tags	> 560 mm	No	
2	0.00%	-11.1%	-28.2%	Tags	> 560 mm	Yes	
3	0.93%	-8.1%	-23.3%	Clips	all	No	
4	0.00%	-18.0%	-33.1%	Clips	all	Yes	
5	0.01%	-9.4%	-24.7%	Clips	> 560 mm	No	
6	0.00%	-11.1%	-28.2%	Clips	> 560 mm	Yes	
7	0.93%	-17.3%	-32.4%	Tags	all	No	all clipped fish are tagged
8	0.00%	-18.0%	-33.1%	Tags	all	Yes	all clipped fish are tagged
9	0.93%	1.9%	2.5%	Clips	all	No	No fish are tagged
10	0.00%	-18.0%	-33.1%	Clips	all	Yes	No fish are tagged
11	0.01%	0.5%	1.3%	Clips	> 560 mm	No	No fish are tagged
12	0.00%	-11.1%	-28.2%	Clips	> 560 mm	Yes	No fish are tagged

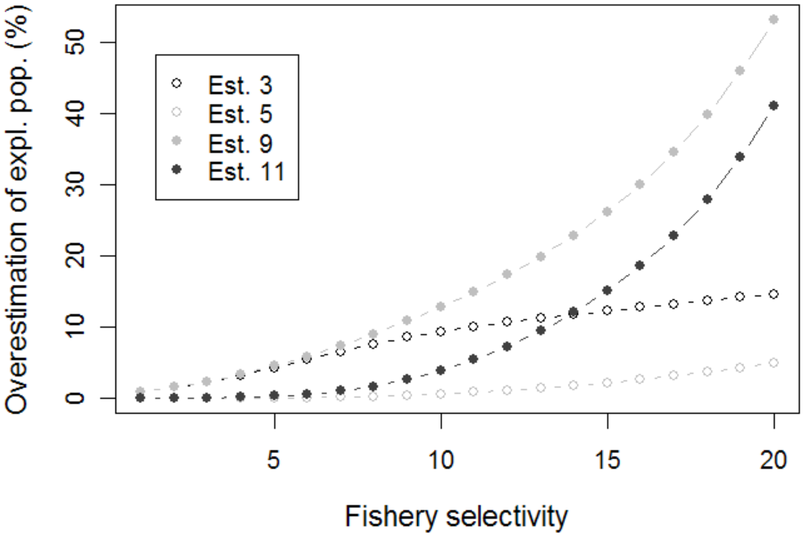
I further examined the sensitivity of all estimators over a wide range of fishery selectivity, varying smoothly from a fishery selectivity that is identical to the 'base' gear selectivity (i.e., no discarding) to a fairly extreme selectivity with 50% selectivity at 620 mm and a slope of 0.02 (Appendix A.17)



Appendix A.17–Range of fishery selectivities used to examine sensitivity of estimator to variability in gear selectivity.

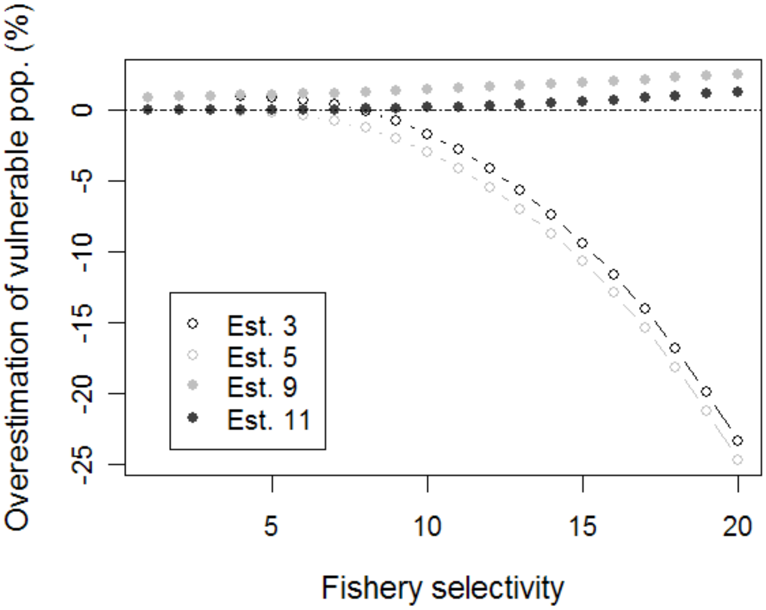
Estimators 9 and 11, which assume that no fish are tagged, are most sensitive to variations in fishery selectivity and can lead to extreme overestimation of exploited population size (over 50% under most extreme scenario). Estimators 3 and, in particular, 5 (Appendix A.5) are much less sensitive to variability in fishery selectivity because the number of clipped fish retained is "inflated" by the higher probability of tagged fish being retained. Even under fairly extreme "highgrading", the assumed higher selectivity of tagged fish (equal to gear selectivity), which increases the number of clips retained, dampens the effect of fishery selectivity on the estimates.

As estimators of vulnerable population size, estimators 9 and 11 are least sensitive to variations in fishery selectivity under the base scenario for pot and gear selectivity (both of which appear to have a strong effect on these results, see Appendix A.11 and Appendix A.16, whereas estimators 3 and 5 are much more sensitive and strongly under-estimate vulnerable population size if fishery selectivity is as pronounced or more pronounced than estimated from NMFS data (the 'base' scenario corresponds approximately to fishery selectivity '14' on the x-axis in Appendix A.18 and Appendix A.19). The under-estimation of vulnerable population size decreases rapidly as fishery selectivity increases because the ratio of retained catch (= "sample size *n*") to retained clips (*m*) decreases as the discarding of small untagged fish (i.e., fishery selectivity) increases.



Appendix A.18–Effect of fishery selectivity on uncorrected estimators (of exploited population size) that are sensitive to variations in fishery selectivity.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.17, from left to right.

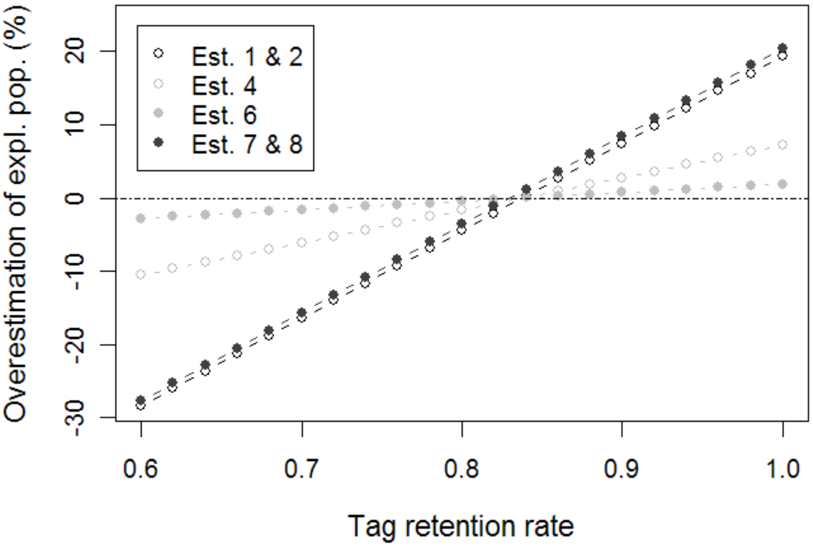


Appendix A.19–Effect of fishery selectivity on uncorrected estimators (of vulnerable population size) that are sensitive to variations in fishery selectivity.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.17, from left to right.

5. Effect of mis-specification of tag retention rates on estimates:

I examined how estimates are affected if the true tag retention rate is not known and is mis-specified in the estimation. Only estimators 1, 2, 4, 6, 7, and 8 are affected by variability in tag retention (other estimators do not depend on tag retention rate). Effects on estimators 1,2, 7 and 8 are nearly identical because all are affected in the same way by a decrease in tag retention, which simply leads to a proportional reduction in the assumed number of marks "available" in the population. Therefore, for example, if tag retention is assumed to be 20% lower than the "true" (but unknown) tag retention rate, these estimates are also 20% lower than the "true" values. In contrast, estimators based on clips are only susceptible to variability in tag retention to the extent that the number of clipped fish retained depends on the number of tagged fish encountered (which have a higher probability of retention than fish with clips only).



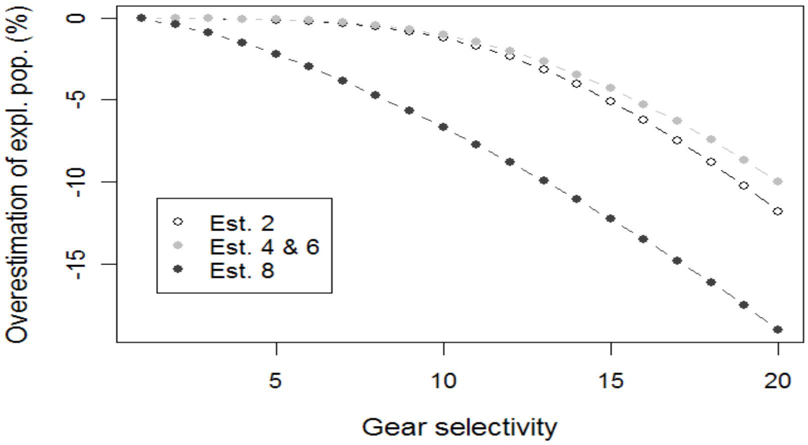
Appendix A.20–Effect of mis-specification of tag retention rate on various estimators. All of the estimators are nearly unbiased at the "true" tag retention rate (0.8376) under the base scenario.

6. Effect of Mis-Specification of Gear Selectivity on Estimates:

I examined sensitivity of estimators of exploited population size to mis-specifying gear selectivity in the estimation. I computed estimates over a range of gear selectivities that varied smoothly between the gear selectivity and the fishery selectivity assumed under the base scenario (as in Appendix A.12). Data used to estimate population size were simulated under the base scenario for pot selectivity, gear selectivity, fishery selectivity, tag retention, and harvest rates.

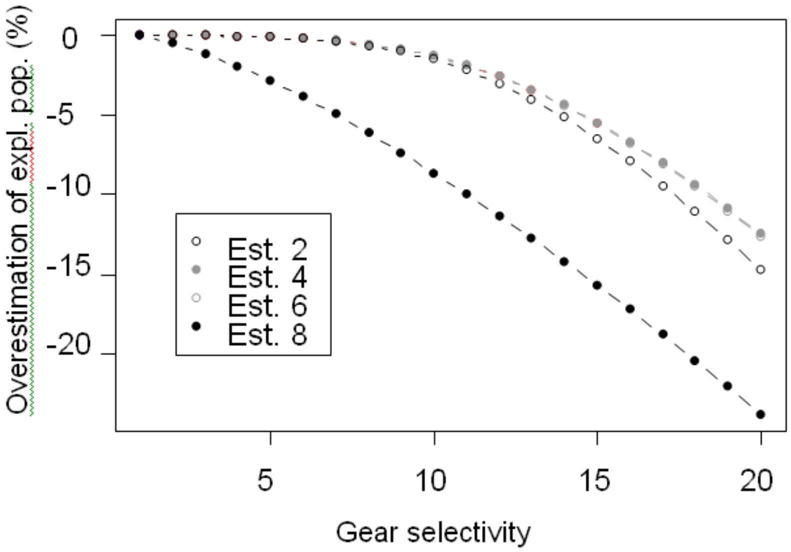
Only those estimators that correct for selectivity and assume that all or large fish are tagged are affected by mis-specification of gear selectivity (2, 4, 6, and 8). Of these, estimators 2, 4, and 6, based on tags (2) or clips (4 and 6), and using either all (4) or only large fish (2 and 6) were affected very similarly under different gear selectivities (Appendix A.21). There is a minimal effect on these estimators if the assumed selectivity is relatively close to the true selectivity, but population size is underestimated up to about 10% as the specified gear selectivity approaches fishery selectivity. Estimator 8, which assumes that all fish are tagged, is much more sensitive to mis-specification of gear selectivity. This is because tags are selected for under the base scenario and if more fish are tagged a larger number of marked fish are retained (higher *m*), leading to a greater underestimation of population size.

Under a scenario with very pronounced dome-shaped selectivity in the pot survey, the relative performance of the estimators is the same as above, but the underestimation is more severe (Appendix A 22). If pot survey selectivity is set to 1 at all size classes (fish greater than 500 mm) the relative patterns of sensitivity of the different estimators are not affected and the magnitude of underestimation of exploited population size is similar to that in Appendix A 21, the base scenario.



Appendix A.21–Effect of mis-specification of gear selectivity on affected estimators of exploited population size under the base scenario.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12 , from left to right, and reflects increasing levels of mis-specification from using the true gear selectivity (1) to using a gear selectivity that is equal to the 'base' fishery selectivity (20).



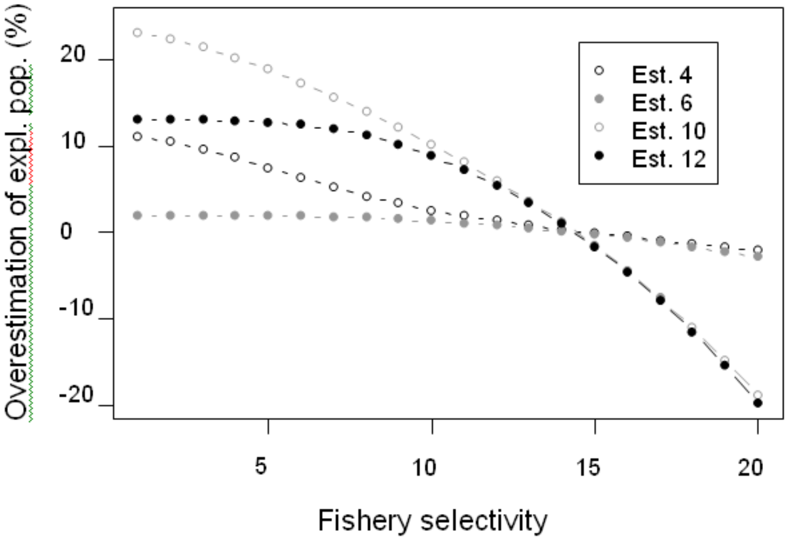
Appendix A.22–Effect of mis-specification of gear selectivity on affected estimators of exploited population size under scenario that assumes extreme dome-shaped selectivity in the pot survey (all other parameters as in base scenario).

7. Effect of mis-specification of fishery selectivity on estimates

I examined sensitivity of estimators of exploited population size to mis-specifying fishery selectivity in the estimation. The true fishery selectivity is essentially unknown for the Chatham Strait fishery, but may be similar to that of federal IFQ longline fishery as assumed under the base scenario. To examine effects of mis-specifying fishery selectivity on the corrected estimators, I computed estimates over a range of fishery selectivities that varied smoothly from a fishery selectivity that is identical to the 'base' gear selectivity (i.e., no discarding) to a fairly extreme selectivity with 50% selectivity at 620 mm and a slope of 0.02 (as in Appendix A.17), which is likely to reflect an upper extreme for gear selectivity. Data used in the estimation were simulated under the base scenario for pot selectivity, gear selectivity, fishery selectivity, tag retention, and harvest rates.

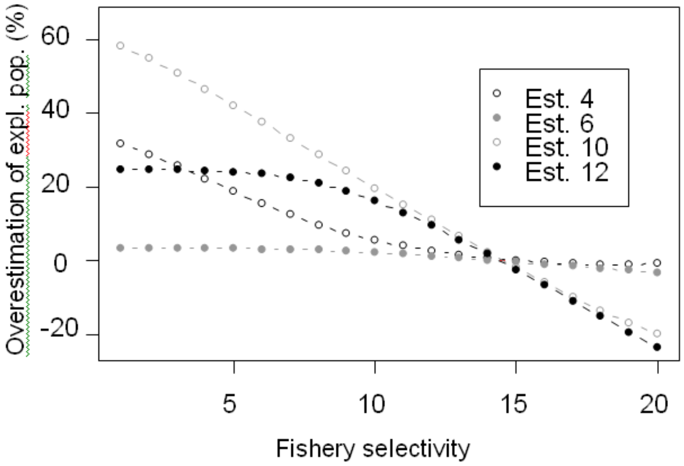
Variability in the specified fishery selectivity affects only those estimators that use a correction factor for fishery selectivity (4, 6, 10, and 12). Estimators 10 and 12 (which assume that no fish are tagged) are more strongly affected by mis-specifications of fishery selectivity than estimators 4 and 6 (Appendix A.23). The latter estimators, under the assumption of full or near-full retention of tagged fish (which are subject to "gear selectivity" under the base scenario), are essentially "buffered" by the high selectivity of tagged fish. This increases the number of clips retained and thereby reduces the influence of the correction factor for fishery selectivity on the estimate (because the correction factor then only applies to clipped fish that were not tagged).

The effects of fishery selectivity were more pronounced if pot selectivity was more dome-shaped (Appendix A.24) or if pot selectivity was equal to 1 at all sizes and the number of small fish, hence the number of small clipped fish, was very high (not shown, similar to Appendix A.24).



Appendix A.23–Effect of mis-specification of fishery selectivity on affected estimators of exploited population size under base scenario.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.17, from left to right.

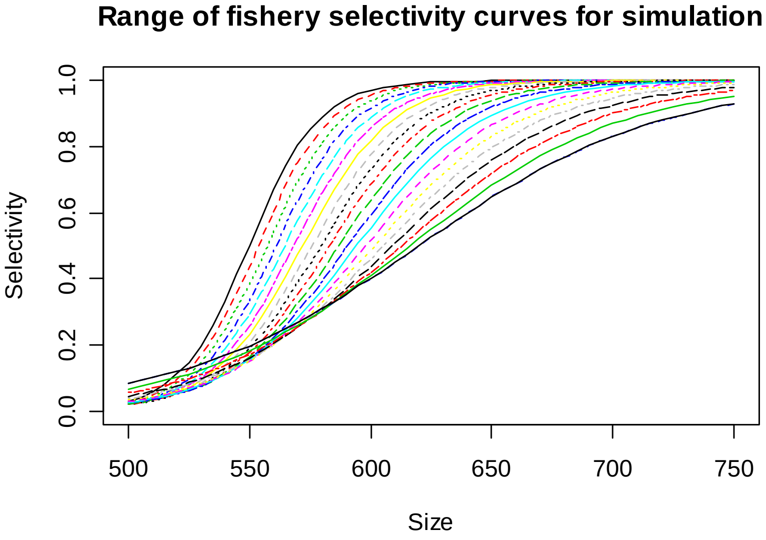


Appendix A.24–Effect of mis-specification of fishery selectivity on affected estimators of exploited population size under scenario that assumes extreme dome-shaped selectivity in the pot survey (all other parameters as in base scenario).

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.17, from left to right.

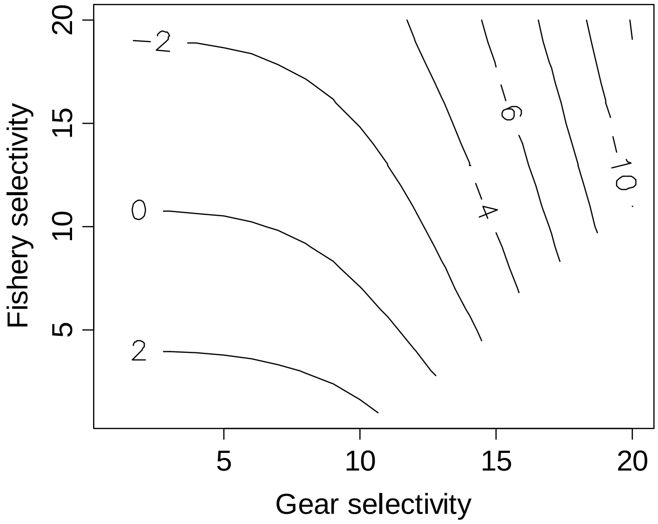
8. Effect of mis-specification of gear selectivity and fishery selectivity on estimates

The final set of simulations examines the effect of mis-specifying both gear selectivity and fishery selectivity on estimates. I used the same range of gear selectivity curves as in Appendix A.12 and the range of fishery selectivity curves shown in Appendix A.25. Effects on the two corrected estimators that depend on both gear and fishery selectivity (4 and 6) was evaluated over a grid of selectivity combinations that paired each of the 20 gear selectivity curves with each of the 20 fishery selectivity curves (Appendix A.26, Appendix A.27).



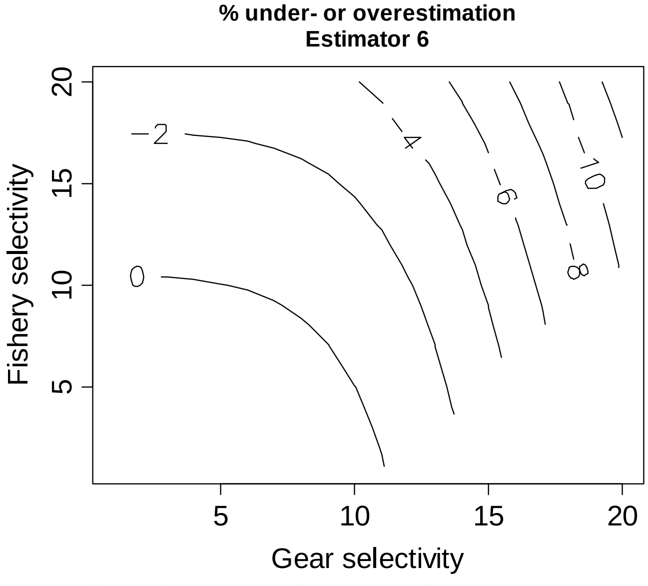
Appendix A.25–Range of fishery selectivity curves used to simulate effects of mis-specifying both gear and fishery selectivity.

Both estimators were relatively robust to mis-specification of selectivities over a fairly wide range of values at which the estimated exploited population size was within ± 2% of the true population size (Appendix A.26, Appendix **A.27**). Therefore, these corrected estimators result in a good estimate of exploited population size even if the true gear or fishery selectivities are poorly known. These results are robust to variability in pot survey selectivity as evident in Appendix A.28 and Appendix A.29, which show the sensitivity to gear and fishery selectivity under relatively extreme scenarios for pot selectivity.



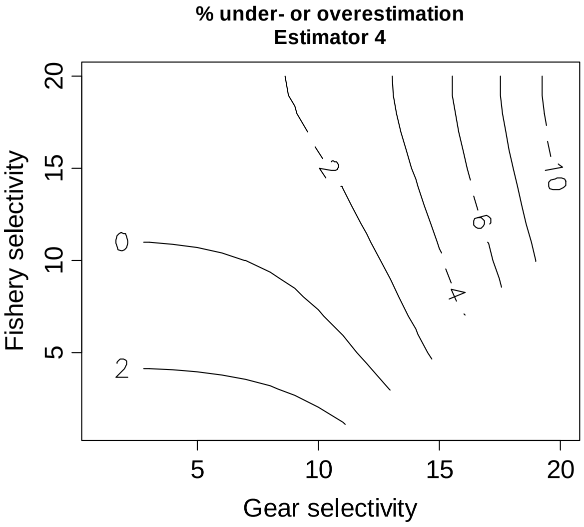
Appendix A.26–Combined effect of mis-specification of gear and fishery selectivity on estimator 4 of exploited population size under base scenario.

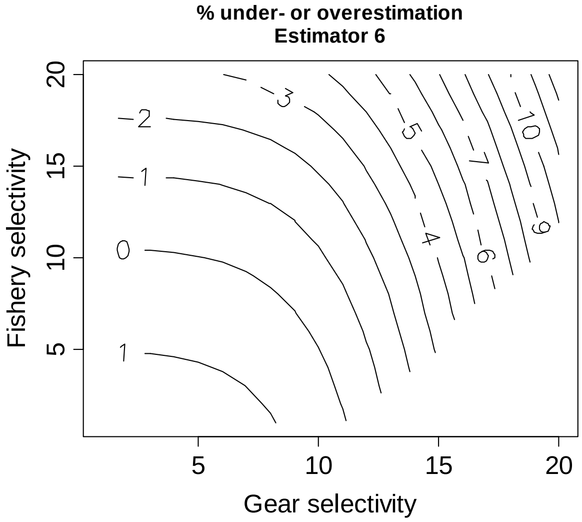
Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12 (gear selectivity) and Appendix A.25 (fishery selectivity). Only scenarios for which 50% fishery selectivity is larger than 50% gear selectivity are shown because fishery selectivity at a given size cannot be less than gear selectivity.



Appendix A.27–Combined effect of mis-specification of fishery selectivity on estimator 6 of exploited population size under base scenario.

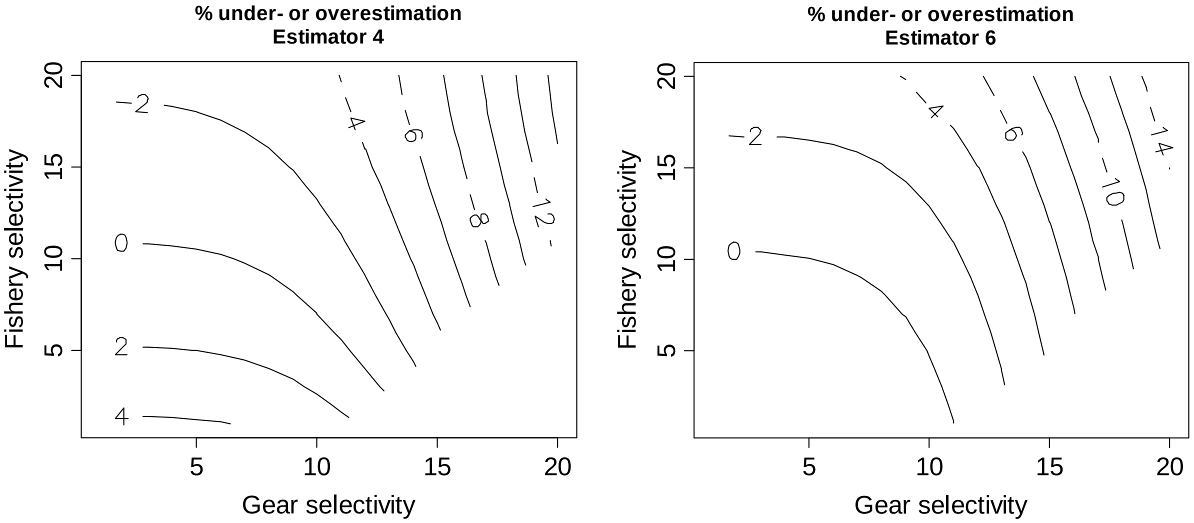
Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12 (gear selectivity) and Appendix A.25 (fishery selectivity). Only scenarios for which 50% fishery selectivity is larger than 50% gear selectivity are shown because fishery selectivity at a given size cannot be less than gear selectivity.





Appendix A.28–Combined effect of mis-specification of gear and fishery selectivity on estimators 4 and 6 of exploited population size under scenario with full pot survey selectivity at all sizes and large number of small fish.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12(gear selectivity) and Appendix A.25 (fishery selectivity). Only scenarios for which 50% fishery selectivity is larger than 50% gear selectivity are shown because fishery selectivity at a given size cannot be less than gear selectivity.



Appendix A.29–Combined effect of mis-specification of gear and fishery selectivity on estimators 4 and 6 of exploited population size under scenario with extremely dome-shaped pot survey selectivity.

Note: Selectivity, from 1 to 20, corresponds to the selectivities shown in Appendix A.12 (gear selectivity) and Appendix A.25 (fishery selectivity). Only scenarios for which 50% fishery selectivity is larger than 50% gear selectivity are shown because fishery selectivity at a given size cannot be less than gear selectivity.

SUMMARY AND CONCLUSIONS

- Size selectivity in the fishery (i.e., in the recapture phase) is important. If not accounted for, exploited population size may be substantially overestimated. If size selectivity is known, suitable correction factors based on the fraction of marks available under a given selectivity pattern can be used to obtain unbiased estimates of exploited population size.
- In general, estimators of the exploited population size of fish greater than 560 mm are much less sensitive to uncertainty about size selectivity because size selectivity is most uncertain at smaller sizes, largely because it is assumed that selectivity is asymptotic and that larger sizes are fully selected by the longline gear. Therefore, estimates based on larger size classes (greater than 560 mm) tend to be more robust in these simulations and management will tend to be more conservative if harvest quotas are based on the abundance of larger fish only. However, better estimates of the true exploited population size, including small fish, will result in more accurate reference points for management (such as a proxy for B_{MSY} and the corresponding fishing mortality).
- The estimators that tend to perform well, particularly those of the population of large fish (greater than 560 mm), are very robust to variability in pot survey selectivity. Therefore, differences in selectivity between the pot survey and the longline fishery are of relatively little consequence in the estimation of exploited population size. Although the pattern of pot survey selectivity is not known, I examined a large range of selectivities including full selectivity at all ages, logistic selectivity, and extreme dome-shaped (or decreasing) selectivity with size (as suggested is the case for the pot survey by additional modeling work). This range is likely to include the true values and the best estimators were robust to this variability.

- Comparison of different estimators suggests that estimators 4 and 6 (corrected estimators based on clips) are most robust to mis-specifications in gear and fishery selectivity (Appendix A.26 to Appendix A.29) and of these two, estimator 6 is less sensitive to mis-specification of the tag retention rate. Therefore, estimator 6 is judged to be the most robust among the estimators examined (under the scenarios examined and given the assumptions). Other estimators of exploited population size suffered from at least the following weaknesses:
    - Estimators 3 and 5 performed relatively poorly over the range of tag retention values examined, even when tag retention was known exactly (Appendix A.11).
    - Estimators 3 and 7 performed poorly under variability in gear selectivity, even if it was known (Appendix A.13).
    - Estimators 9 and 11 performed poorly under variability in fishery selectivity, even if it was known (Appendix A.18).
    - Estimator 3 performed relatively poorly under variability in fishery selectivity, even if it was known (Appendix A.18).
    - Estimators 1,2,7,8 performed poorly if tag retention was mis-specified (i.e., different from the true tag retention rate) (Appendix A.20).
    - Estimator 8 performed poorly if gear selectivity was mis-specified (Appendix A.21, Appendix A.22).
    - Estimators 10 and 12 performed poorly if fishery selectivity was mis-specified (Appendix A.23, Appendix A.24).
    - Estimator 4 performed poorly if fishery selectivity was substantially under-estimated (Appendix A.23, Appendix A.24).
    - By process of elimination I concluded that estimator 6 was the best estimator: it is only weakly affected by variability in or mis-specification of tag retention, corrects for gear and fishery selectivity when they are known, and is relatively robust to mis-specification of gear and fishery selectivity if they are not accurately known. However, the estimator provides an estimate of the population above 560 mm only, which assumes that the number of exploitable smaller fish (less than 560 mm) can be ignored. Estimator 4 performed equally well under most scenarios but was somewhat more sensitive if the true fishery selectivity is underestimated in computing the correction factor. It has the advantage of providing an estimate of total exploitable biomass.

MATURITY AT AGE

For most age-structured analyses, as well as for yield-per-recruit analyses, maturity-at-age schedules are used to estimate the total or female spawning-stock biomass in a given year. Maturity at age is used because the models are age-structured rather than size-structured, although maturity for many species may be a function of size rather than age. Most stock assessments that I am familiar with use maturity at age. One exception was the Pacific cod stock in the Bering Sea, which uses a length-based assessment because of uncertainties about aging cod and had used a maturity-at-length schedule in the assessment for many years. However, the most recent assessment for cod used an age-based schedule.

While maturity at age is likely to vary over time with potentially important implications for stock assessment, all of the stock assessments in the Bering Sea and Gulf of Alaska currently use a fixed maturity-at-age schedule. This is primarily a data limitation, because few surveys are done that sample fish close to spawning and the number of fish examined for maturity is quite small for most species. To my knowledge, walleye pollock in the Gulf of Alaska is the only stock that has a reasonable time series of maturity at age, but even for this stock, the assessment uses a fixed maturity-at-age schedule based on data from winter surveys conducted 1983–2002.

There are advantages to using a fixed maturity-at-age schedule, which provides consistency over time and avoids large variability in the estimates of spawning stock over time that would arise from sampling variability (i.e., if maturity at age is updated annually). For example, although a re-evaluation of maturity of pollock in the Bering Sea found some interannual variability (Stahl 2004), the pollock stock assessment continues to use a fixed maturity-at-age schedule based on Smith (1981), which was reasonably consistent with maturity at length as determined by Stahl (2004).

Gulf of Alaska Sablefish

The NOAA stock assessment for Gulf of Alaska sablefish uses maturity at age estimated from a logistic regression fit to data in Sasaki (1985, cited in Hanselman et al. 2006). Until 2005, the assessment used average male and female maturity from this study to compute spawning biomass and has used female maturity at age since then (Hanselman et al. 2006). The NOAA is currently re-evaluating maturity at age to update the maturity schedule.

Chatham Strait Sablefish

I examined both fishery and longline survey data from Chatham Strait to estimate maturity at age using a linear logistic regression model. The model uses the logit-transformation to fit a simple linear regression of the probability of being mature on age:

$$\mathrm{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta \cdot age_i$$

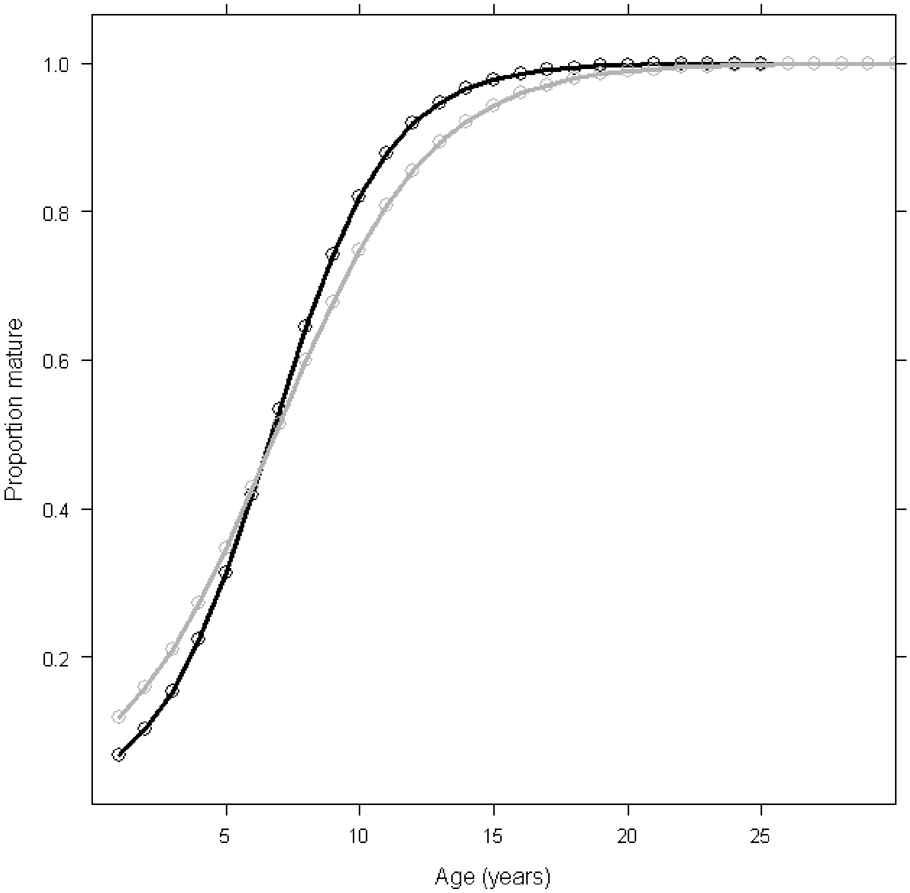
where p_i is the probability that the i^{th} individual is mature and α and β are regression parameters. Maturity is coded as 0 (maturity codes 1 or 2) or 1 (maturity code 3 or higher) and the model assumes that maturity follows a binomial distribution. I fit the model by maximizing the binomial likelihood using function ‘glm’ in R. When the equation is solved for p_i, it can be written in the familiar logistic equation:

$$p_i = \frac{e^{\alpha+\beta \cdot age_i}}{1+e^{\alpha+\beta \cdot age_i}} = \frac{1}{1+1/\left(e^{\alpha+\beta \cdot age_i}\right)} = \frac{1}{1+e^{-\beta(age_i - a50)}}$$

where *a50* = α/β is the age at 50% maturity.

When longline survey data from all years (1988–2006) are combined, the proportion mature at age differs significantly between males and females (analysis of deviance: p = 0.009, Appendix B.2) with ages at 50% maturity of 6.35 and 6.15 for females and males, respectively. I fit the model separately by year, although sample sizes are relatively small in many of the early years as shown in this table:

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
N	146	110	81	117	104	180	160	86	157	261	192	115	213	181	301	388	381	456	417



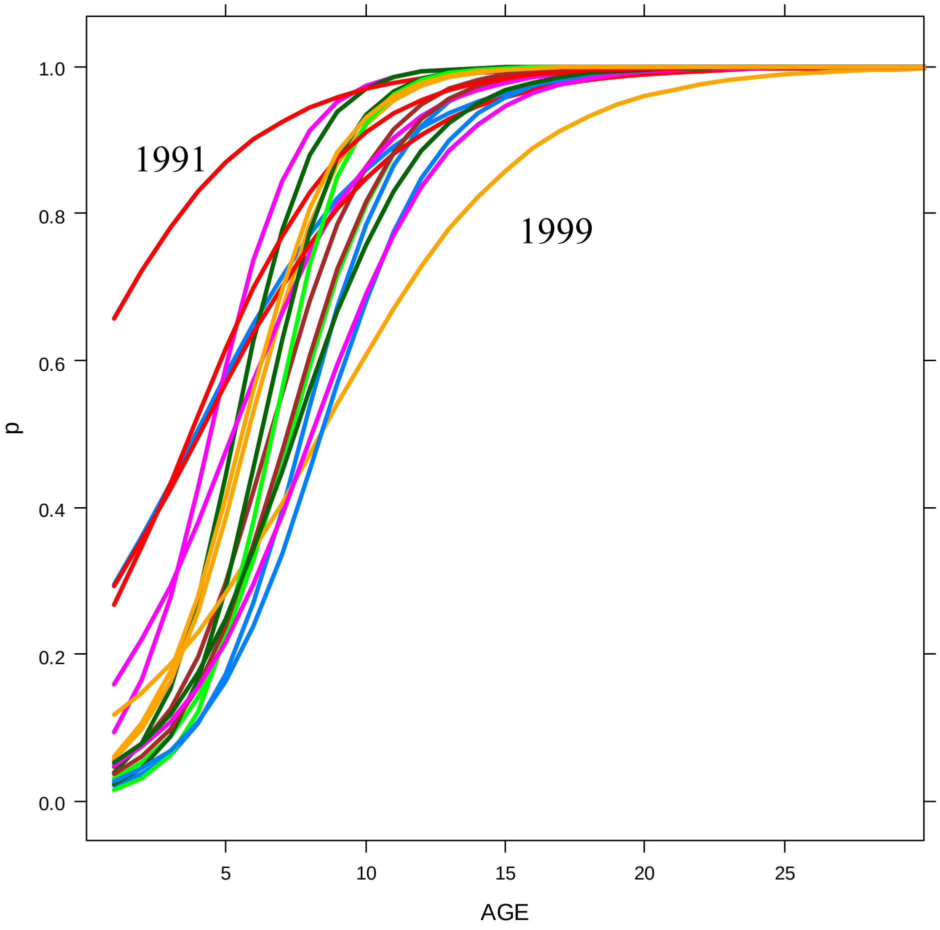
Appendix B.2–Estimated proportion mature at age by sex (black = females, grey = males) based on longline survey data, 1988–2006 (all years combined).

Nevertheless, the model shows significant differences in maturity at age among years (Appendix B.3, Appendix B.4, p < 0.001). In particular, the years 1991 and 1999 stand out as having a larger proportion of mature fish by age. Maturity at age has been relatively consistent since 2000, therefore I fit the logistic regression model to the 2000–2006 data from the longline survey for use in the age-structured assessment, after eliminating two outliers (23/32-year old females classified as immature). The estimated maturity-at-age schedule suggests that in recent years 50% maturity was reached at age 6.88, which compares well with the maturity schedule used in the NOAA assessment, which estimates 50% female maturity at age 6.60 (Appendix B.5). However, the NOAAS fit rises much more steeply with age. At least some of the difference may be due to differences in aging between NOAA and ADF&G age readers.

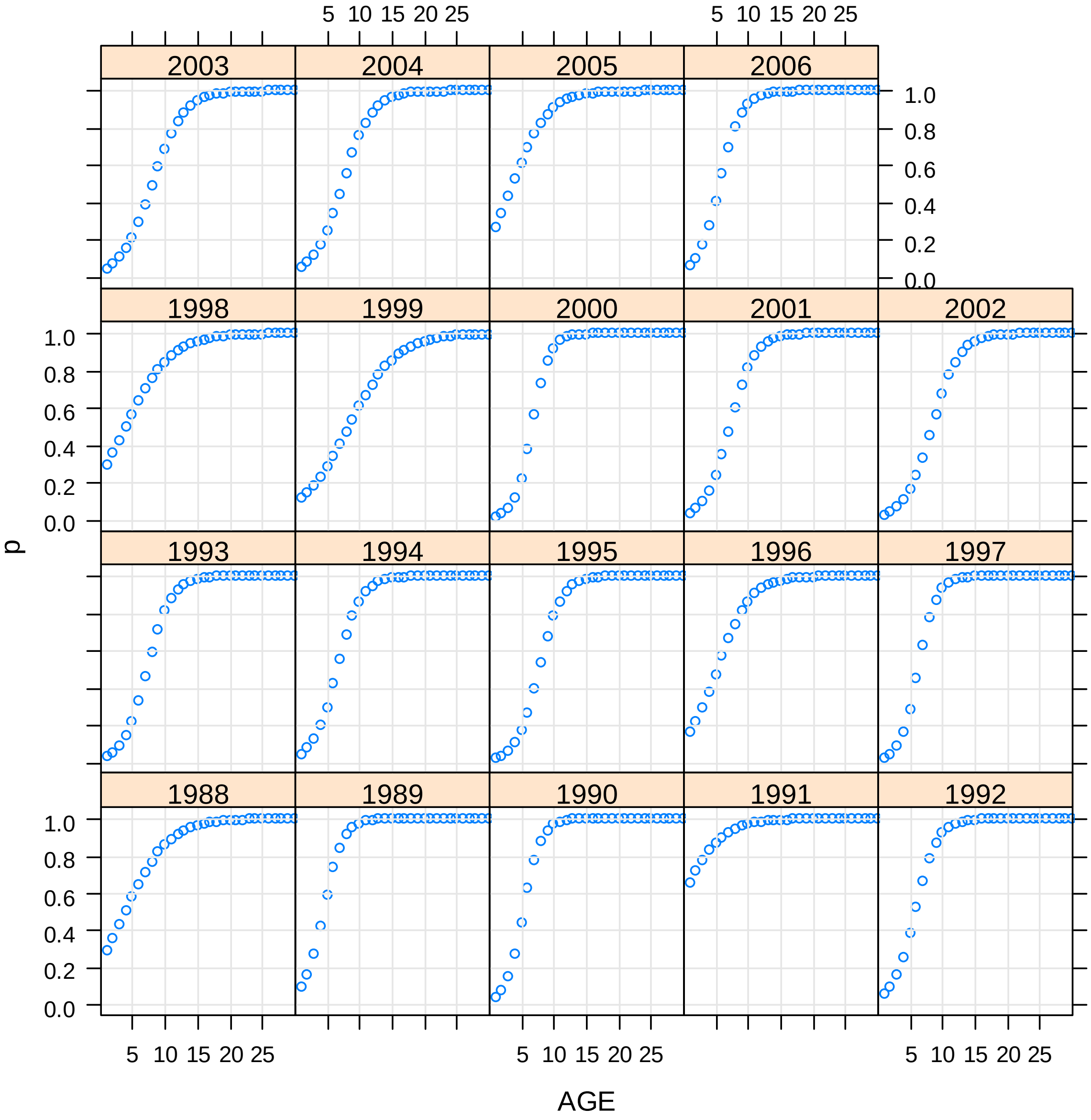
The estimated proportions mature at age are:

Age	2	3	4	5	6	7	8	9	10	11	12	13	14
Prop.	7.22	11.61	18.14	27.21	38.67	51.55	64.22	75.18	83.63	89.61	93.57	96.08	97.64
Age	15	16	17	18	19	20	21	22	23	24	25	26	
Prop.	98.59	99.16	99.50	99.70	99.82	99.90	99.94	99.96	99.98	99.99	99.99	100.00	

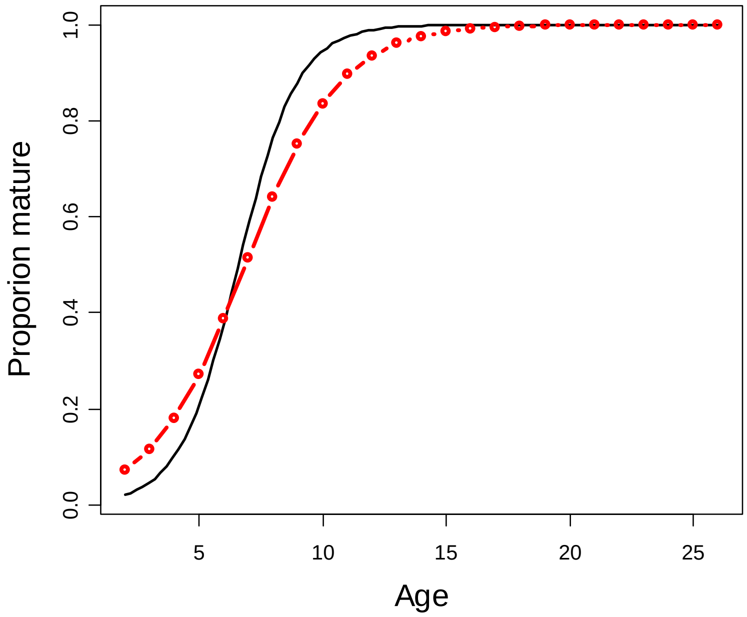
I also compared estimates from the fishery to those from the longline survey. As might be expected due to selection of larger fish in the fishery, there is a substantial difference in maturity-at-age between estimates based on the fishery and estimates based on the longline survey (Appendix B.6).



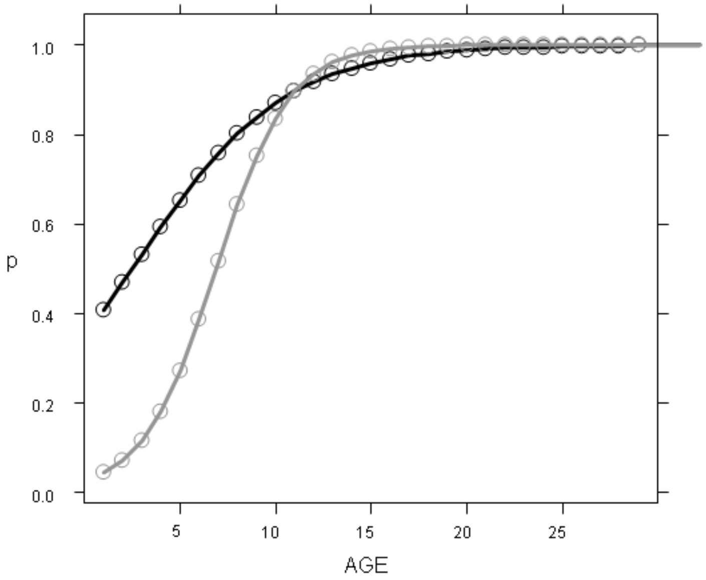
Appendix B.3–Proportion of female sablefish mature (p) by age and year, 1988–2006, estimated from longline survey data.



Appendix B.4–Proportion of female sablefish mature (p) by age and year, 1988–2006, estimated from longline survey data.



Appendix B.5–Estimated proportion of female sablefish mature based on 2000–2006 longline survey data (dashed line) compared to logistic fit used in NOAA sablefish assessment (solid line).



Appendix B.6–Estimated proportion of female sablefish mature (p) based on 2000–2006 longline survey data (grey lines) and on 2000–2006 fishery data (black line).

SEX RATIOS

To estimate the proportion of females in the population, in order to compute accurate estimates of spawning biomass, I examined trends in the ratio of females to males in the longline survey data and in the fishery data. I used biological data for sablefish from all random samples, assuming that the aged fish in these samples are representative of the exploitable population. Similarly, I used random samples from the fishery age-weight-length data.

Average Sex Ratios Over Time

The proportion of females in the longline survey over time fluctuates from 38% (1995) to 56% (in 2006), with an overall average ratio of close to 50:50 (49.8% females) (Appendix B.7). The proportion of females in fishery catches was higher than in the longline survey, ranging from 55% to 68% between 2000 and 2006 with an overall average of 61% (Appendix B.8). In both the survey and the fishery, the proportion of females appears to be increasing in recent years.

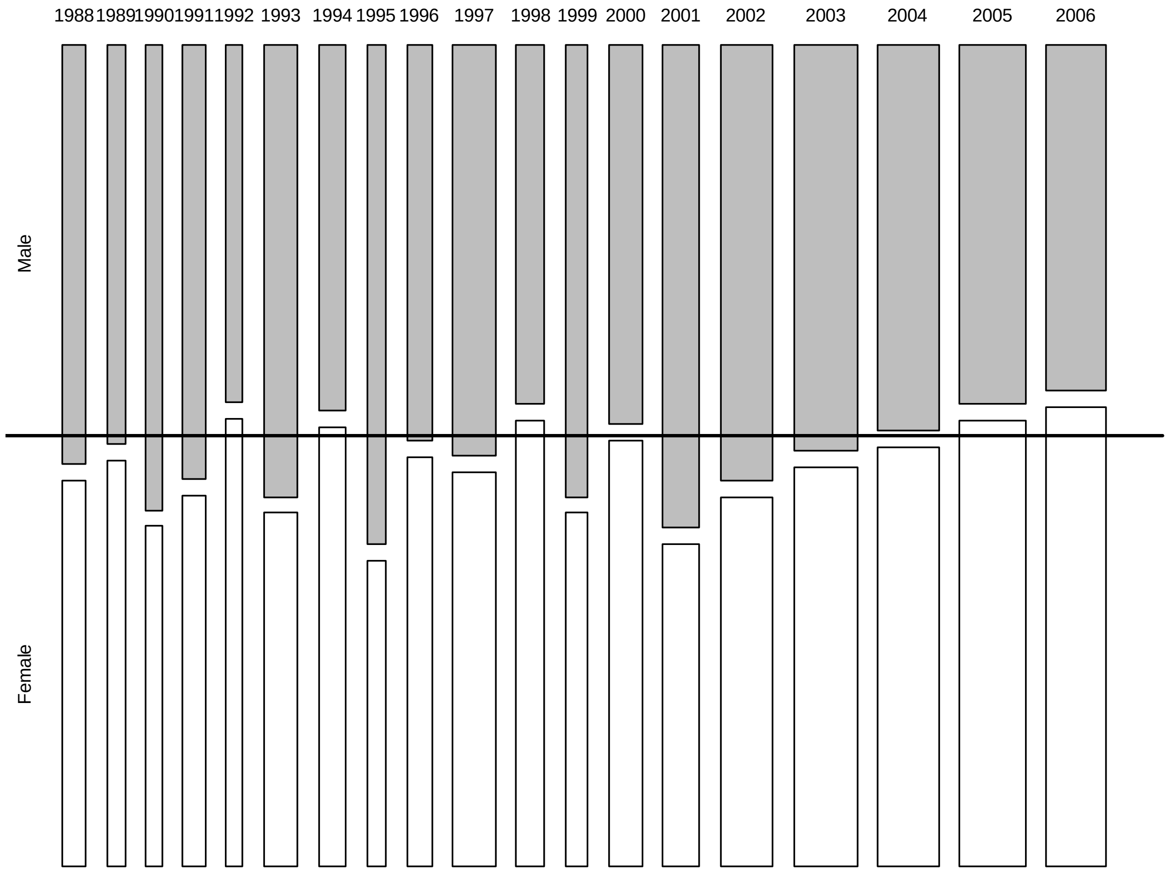
Average Sex Ratios by Age

There was a clear tendency in both the longline survey samples (Appendix B.9) and in the fishery samples (Appendix B.10) for the proportion of females to decrease at older ages. The trend was parallel in the survey and in the fishery, but the average proportion of females was generally higher at a given age in the fishery samples, which may reflect size selectivity in the fishery or reduced vulnerability of males to the longline gear at the time of the fishery. Reduced vulnerability could be due to, for example, reduced feeding by males in late fall or males leaving the area. The decrease in the proportion of females with age could be a result of differences in vulnerability between sexes at older ages, a higher natural mortality rate of females, and/or a higher fishing mortality of older females due to their larger size-at-age.

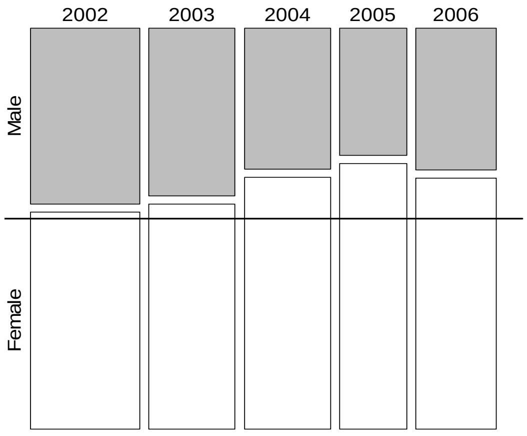
I examined whether differences in the sex ratio overall could be attributed primarily to age or to variability in sex ratios over time. I modeled the proportion of females from the longline survey as a function of age and year using a logistic regression. It appears that most of the variability, by far, can be attributed to age (Appendix B.11). I therefore ignored variability over time and estimated the proportion of females by age class across all years (Figure 52):

Age:	2	3	4	5	6	7	8	9	10	11	12	13	14
Prop:	54.4	54.7	54.9	55	55.1	55.2	55.1	54.9	54.6	54.1	53.4	52.5	51.5
Age:	15	16	17	18	19	20	21	22	23	24	25	26	
Prop:	50.3	49.2	48	46.9	45.8	44.7	43.6	42.6	41.6	40.6	39.8	34.2	

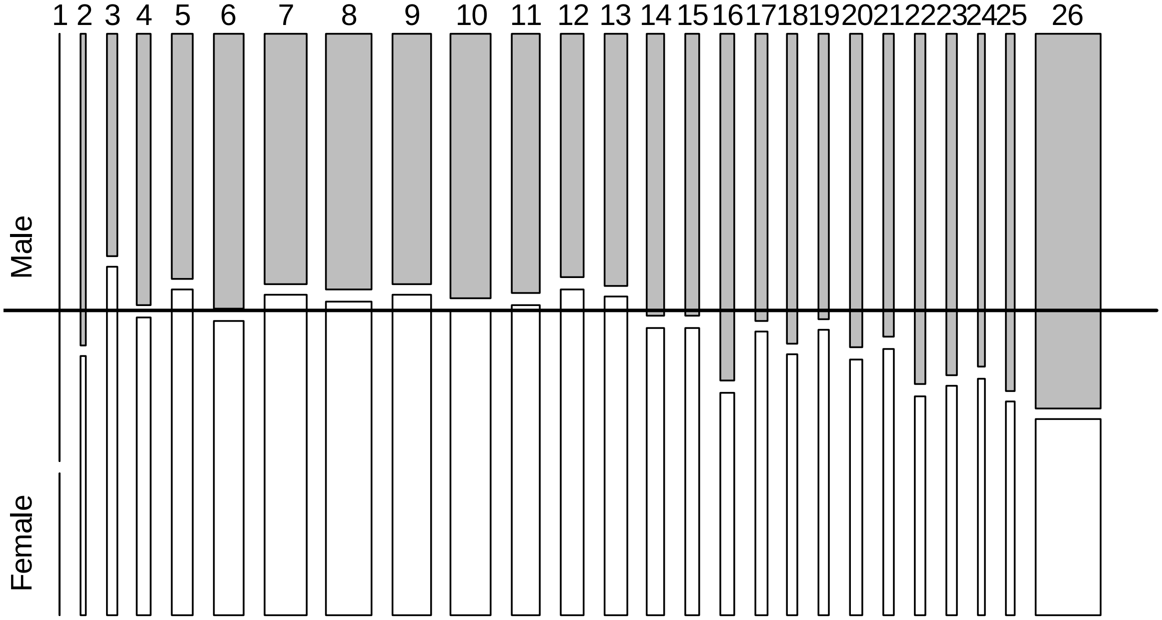
These proportions are used in the current version of the age-structured model to estimate spawning biomass. It may be possible to use higher natural and fishing mortality for females in a split-sex model to see if differential mortality can account for the changing sex ratio with age. In that case the proportion of females by age would be implicitly estimated in the model and could even be fit to observed proportions.



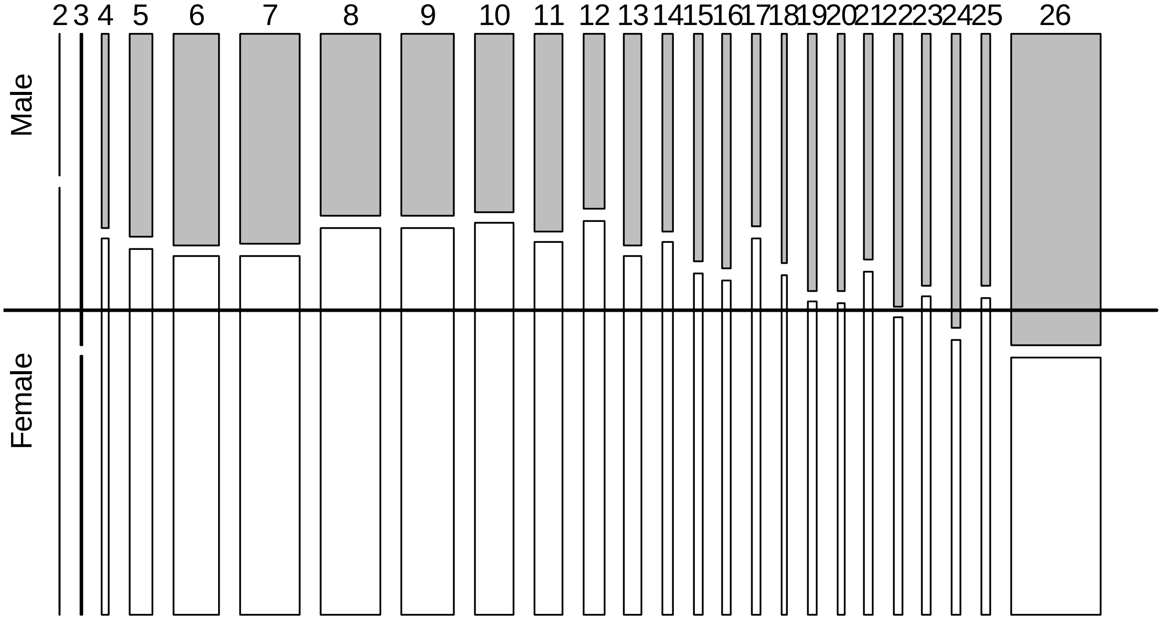
Appendix B.7–Proportion of females and males in random samples from longline survey, 1988–2006. Horizontal black line represents 50% male and female. Width of bars is proportional to number of samples.



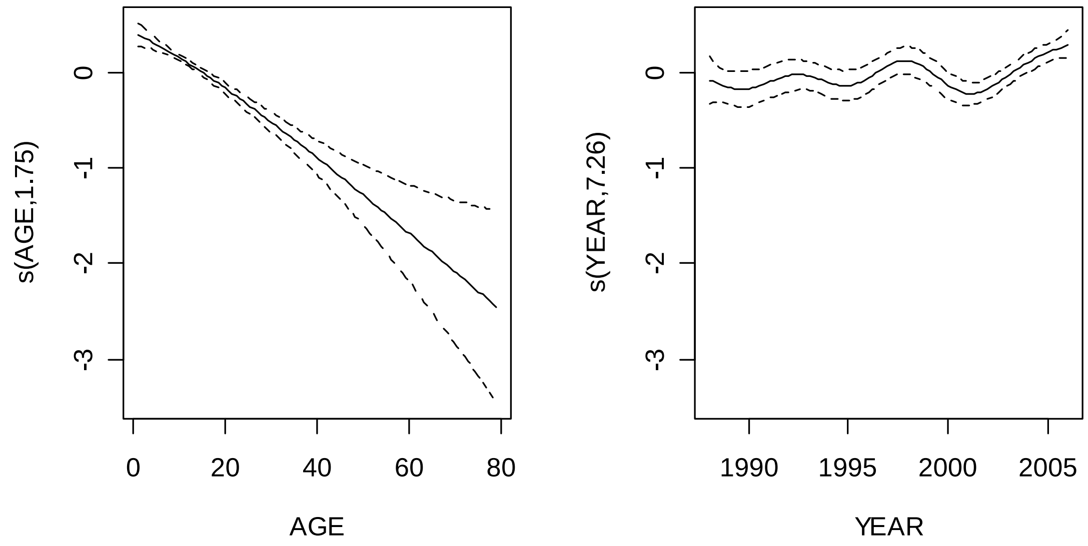
Appendix B.8–Proportion of females and males in random fishery samples, 2000–2006. Horizontal black line represents 50% male and female. Width of bars is proportional to number of samples.



Appendix B.9–Proportion of males and females in random samples from longline survey by age. Horizontal black line represents 50% male and female. Width of bars is proportional to number of samples in each age class. Age 26 denotes ages 26 and older.



Appendix B.10–Proportion of males and females in random fishery samples by age. Horizontal black line represents 50% male and female. Width of bars is proportional to number of samples in each age class. Age 26 denotes ages 26 and older.



Appendix B.11–Estimated proportion of females (on logit scale and normalized) as smooth functions of age and year (smooth time trend, similar results if separate proportion is estimated for each year).