ISSN: 0074-7246

INTERNATIONAL PACIFIC HALIBUT COMMISSION

ESTABLISHED BY A CONVENTION BETWEEN
CANADA AND THE UNITED STATES OF AMERICA

Scientific Report No. 83

Assessment and management of Pacific halibut: data, methods, and policy

by

William G. Clark and Steven R. Hare

SEATTLE, WASHINGTON
2006

The International Pacific Halibut Commission has three publications: Annual Reports (U.S. 0074-7238), Scientific Reports, and Technical Reports (U.S. ISSN 0579-3920). Until 1969, only one series was published (U.S. ISSN 0074-7246). The numbering of the original series has been continued with the Scientific Reports.

INTERNATIONAL PACIFIC HALIBUT COMMISSION
P.O. BOX 95009
SEATTLE, WASHINGTON 98145-2009, U.S.A.
www.iphc.washington.edu

Assessment and management of Pacific halibut: data, methods, and policy

Contents

Abstract

The International Pacific Halibut Commission (IPHC) sets annual catch limits by regulatory area for the directed halibut fisheries in the northeast Pacific Ocean. Abundance in each area is estimated by fitting an age- and sex-structured population model to commercial and survey data. A biological target for total removals, called the "constant exploitation yield" (CEY), is then calculated by applying a carefully chosen target harvest rate to the estimated exploitable biomass in each area. The catch limits recommended by the staff to the Commission may be somewhat higher or lower than the CEY depending on a number of technical and policy considerations. The Commissioners make the final decision. This paper details the data and the model used in the annual stock assessment (exemplified by the 2004 assessment), summarizes our present understanding of stock dynamics, and describes the constant harvest rate policy.

Assessment and management of Pacific halibut: data, methods, and policy

William G. Clark and Steven R. Hare

Overview

The aim of this paper is to document the data and methods used by the staff of the International Pacific Halibut Commission (IPHC) in its annual assessment of the abundance and potential yield of the stocks. The focus is on how stock size and productivity are estimated rather than on a particular set of estimates. The 2004 assessment model is described in detail, but only for the purpose of showing the quality and retrospective behavior of some actual fits. There are usually a few minor changes in model parameterization or weighting every year, along with an extra year of data, and the estimates of abundance change from year to year as a result. The most recent annual stock assessment document (posted on the IPHC website) should be consulted for the exact form of the latest model and the latest estimates of stock status.

There are actually two components of the stock assessment. Present abundance is estimated by fitting a modern age- and sex-structured model to survey and commercial data from recent years covered by setline surveys, meaning back to 1974 for the eastern Gulf of Alaska (westward to Kodiak Island) and back to 1996 for areas farther west. In a second step, abundance in the eastern Gulf is estimated back to 1935 by beginning with the modern estimates for 1974 and fitting a simpler model to commercial data for earlier years. We cannot estimate historical abundance for areas west of Kodiak Island because those areas were only lightly fished until the mid-1990s.

In what follows, the first two introductory sections summarize halibut distribution and life history, and the development of the fishery and Commission management. The next two sections describe the various kinds of data used in the modern assessment and how they are compiled and preprocessed external to the assessment model. The next two sections set out the structure of the modern assessment model, how it is fitted to the data, and how well it performs in terms of goodness of fit, retrospective behavior, and variance of the abundance estimates. The next section details the historical model used to estimate abundance in years before 1974. The full series of abundance estimates (and growth data) for the eastern Gulf of Alaska from 1935 to the present are the basis of our analysis of population dynamics, summarized in the penultimate section. These dynamic relationships are used in simulations that guide our choice of a target constant harvest rate and other elements of the Commission's harvest policy, as described in the last section.

Biological background

Pacific halibut (*Hippoglossus stenolepis*) are widely distributed in coastal waters of the northeast Pacific from central California around the Gulf of Alaska out the Aleutian Island chain and into the Bering Sea, with a center of abundance around Kodiak Island (Fig.1). About 2% of the biomass is off Oregon and Washington, about 15% off British Columbia, and the remainder off Alaska. The species also occurs on the Asian side, but this paper deals only with North American waters where halibut are studied and managed by the International Pacific Halibut Commission.

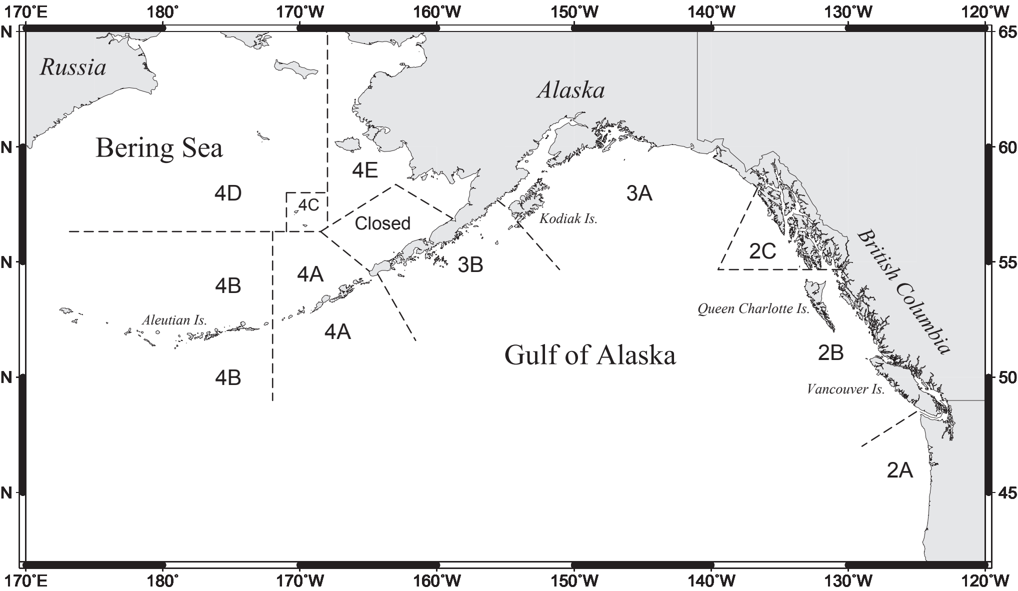


Figure 1. IPHC regulatory areas.

In summer halibut are distributed on the continental shelf and upper slope. In winter mature fish migrate to spawning grounds deeper on the slope (IPHC 1998). The eggs, larvae, and postlarvae drift in the currents for about six months before settling out and metamorphosing to the flatfish form. This drift generally transports fish spawned in the Gulf of Alaska westward, with the result that the major nursery grounds are in the western Gulf and Bering Sea, and few very young fish are found east of Kodiak Island. Stocks in the eastern Gulf of Alaska are replenished by juvenile fish that migrate eastward from the western nursery grounds (Skud 1977, St-Pierre 1989, Clark and Hare 1998). Recoveries of fish marked in summer at lengths over 65 cm (6-7 years old) are mostly made near the release location, indicating that by that age fish have completed the migration from the nursery grounds and thereafter occupy the same summer feeding ground year after year (Trumble et al. 1990). Recoveries of mature fish marked in winter are often made some distance away in summer, however, showing that some fish undertake a substantial spawning migration in winter. The fish off Oregon, Washington, and most of British Columbia in particular must migrate north to spawn because there are no significant spawning grounds south of the Queen Charlotte Islands (St-Pierre 1984).

Genetic studies in the past using protein electrophoresis have shown differences between halibut stocks on the eastern and western sides of the North Pacific, and also between Atlantic halibut (*Hippoglossus hippoglossus*) and Pacific halibut, but no differences within the northeast Pacific (Grant et al. 1984). Further research on this question is being conducted with modern methods (Hauser et al. 2006). For the time being we regard the halibut in the northeast Pacific as a single spawning stock. Separate catch limits are set for each of the regulatory areas shown in Figure 1, and we sometimes speak of e.g. the "Area 3A stock", but these are management stocks rather than biological stocks. The setting of catch limits by regulatory area serves to allocate the harvest among areas (and between the United States and Canada) more or less in proportion to abundance.

Female and male halibut both grow to a length of about 60 cm at age 6. Thereafter females grow faster and reach substantially greater sizes. All really large halibut are females. The modal length in commercial landings is around 100 cm. As explained below in the section on stock dynamics, growth rates of both sexes have varied greatly over the last century (Clark and Hare 2002). Females reach sexual maturity at an average age of 11 years, males somewhat earlier. A few fish older than 40 years have been observed in samples, but over 90% of the commercial catch consists of fish 7-20 years old.

Fishery and management background

Aboriginal peoples in North America have fished halibut for thousands of years. Commercial longline fisheries based in Seattle and Vancouver developed shortly after the completion of the first transcontinental railroads to those cities late in the nineteenth century. In the early years of the twentieth century the fishery went through the classic boom-and-bust cycle. Fishing effort and catches increased rapidly at first, then catch rates dropped, and eventually the total catch peaked and declined as well, despite the continuing increase in fishing effort (Thompson and Freeman 1930, Bell 1981). The industry in both countries petitioned the governments for relief, and in 1923 they signed a convention establishing the International Fisheries Commission (renamed the International Pacific Halibut Commission in 1953) to conduct biological studies and recommend management measures. The first Director of the Commission, W. F. Thompson, was a giant in the history of fishery science who in a few short years both collected the data and developed the methods needed to assess the stock and determine what level of catches could be sustained (Thompson and Bell 1934). He recommended to the governments that the Commission be authorized to define regulatory areas, set catch limits, and adopt other regulations. The governments gave their assent in a new convention signed in 1930, and the Commission commenced quota management in 1932.

The stock and the fishery recovered under Commission management in the 1930s and fared well until the 1960s under the operation of Thompson's principles (Fig. 2). During the 1960s distant-water trawl fleets arrived in the northeast Pacific and took a large bycatch of halibut (Williams et al. 1989). Recruitment to the halibut stock in these years was poor (very possibly because of the trawl bycatch of juveniles). The Commission was slow to reduce catch limits in the directed longline fishery because under current international agreements the coastal states were obliged to demonstrate "full utilization" of the halibut stock to ward off a directed distant-water fishery. As a result of all these developments the stock declined steeply during the 1960s and by the early 1970s had fallen back to the low level reached previously in the early 1930s. Faced with this crisis, the Commission acted resolutely. Catch limits were drastically reduced and for a decade were kept below the estimated surplus production in order to rebuild the stock.

Once again the lowered catch limits were effective. Thanks in part to a regime shift in the climate of the North Pacific in 1977 that approximately doubled recruitment (Clark and Hare 2002), the stock rebounded in the late 1970s and early 1980s. In 1984 the Commission declared the stock rebuilt and adopted a constant harvest rate policy for setting catch limits, which has continued to the present. Each year the staff estimates abundance in each regulatory area by fitting a population model to commercial and survey data going back to 1974. A biological target level for total removals in each area is calculated by applying a carefully chosen target harvest rate to the estimate of exploitable biomass. This biological target level is called the "constant exploitation yield" or CEY. Part of the total yield is set aside to provide for miscellaneous removals (e.g., bycatch in other fisheries, sport and subsistence catches in Alaska). The remainder is available for directed fisheries subject to allocation, which are the commercial longline fisheries in all areas and the sport fisheries in Areas 2A and 2B. This amount is called the "fishery CEY." Staff catch limit recommendations may be lower or higher than the calculated fishery CEY depending on the Director's assessment of the uncertainties and risks involved in each regulatory area. The Commissioners make the final decision. at the annual meeting in January after considering the recommendations of the staff, the industry and the two governments' scientific advisers.

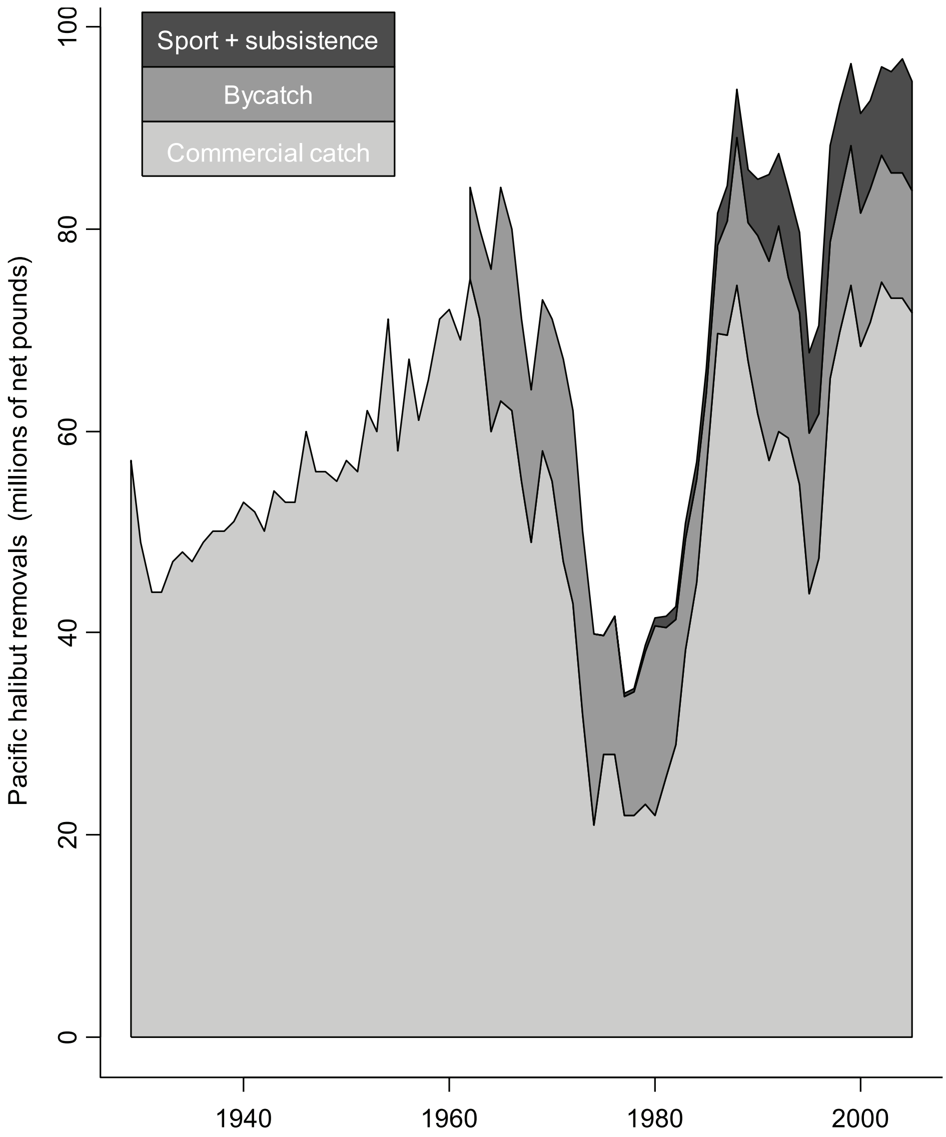


Figure 2. Removals of Pacific halibut from all IPHC areas, 1929-2005. The bycatch figures refer to bycatch mortality and include sublegals. There are no estimates of bycatch before 1962; it was probably a few million pounds per year in the late 1950s and negligible before 1955.

During most of the twentieth century Canadian and U.S. halibut boats fished coastwide, and about half the catch in Alaska waters was taken by Canadian vessels. When both countries extended their maritime jurisdiction in 1976, Canadian vessels were expelled from U.S. waters and vice versa. Canada carried out a buyback program to reduce its fleet to a size more appropriate for the yield available from Canadian waters, but it was still larger than needed. In Alaska a flood of new vessels entered the fishery. In both countries the fishing seasons grew shorter and shorter during the 1980s, to only a few days in most areas. These years are remembered as the "derby fishery" and not fondly. The fishery was hectic, chaotic, and dangerous, and fish quality suffered. Canada adopted an individual quota system in 1991 which eliminated the problems associated with the derby fishery and allowed vessel owners to fish more efficiently and profitably. Alaskans who had adamantly opposed individual quotas were quickly won over by the Canadian example, and an individual quota system was adopted for Alaska in 1995. Today the derby fishery survives only in Washington and Oregon.

The bycatch of halibut by distant-water fleets was brought under strict control after Canada and the United States extended jurisdiction in 1976, but bycatch has remained a contentious issue between the halibut fishery and other domestic fisheries, particularly the groundfish trawl fisheries. It has also been a contentious issue between Canada and the United States because the large trawl bycatch of juvenile halibut in Alaska (mainly in the Bering Sea) must include some fish that would otherwise migrate to Canada and recruit to the fishable stock there. Lacking detailed knowledge of juvenile distribution and migration, it is not possible to make good estimates of the area-specific impacts of the bycatch of different sizes of halibut in different parts of Alaska. Simulation studies using a range of assumptions indicate that the impact falls mostly but not entirely in the area where the bycatch is taken. At present there is a two-part process for dealing with bycatch in calculating fishery CEY. The bycatch of fish above the commercial minimum size limit (81 cm), which have presumably completed their juvenile migration, is deducted from the total CEY in the regulatory area where they are caught. The coastwide recruitment loss resulting from sublegal bycatch—estimated to be about 10%—is included in the simulations that are conducted to choose a target harvest rate. It therefore depresses the target harvest rate slightly in all areas, but the choice of an optimum harvest rate is not at all sensitive to this factor. This method of accounting for juvenile bycatch therefore finesses the uncertainty about unequal and unknown area-specific impacts of juvenile bycatch (Clark and Hare 1998).

Assessment data

The annual stock assessment uses data from commercial landing reports, commercial logbooks, port sampling of commercial landings, IPHC setline surveys, and fishery agencies in both countries that report estimates of bycatch, sport catch, and subsistence catch. This section describes each data type.

Commercial fishery data

The weight of every commercial landing is recorded on a sales report (fish ticket), a copy of which is sent to the IPHC. The total catch in weight in every regulatory area in every year is known from this reporting system. The weight reported is net weight, meaning headed and gutted weight which is about 75% of round weight. Curiously, this measure of weight is used throughout in halibut assessment and management, so for example estimates of biomass in the sea are stated in net weight not round weight. In 2004 commercial landings totaled 73 million (net) pounds.

IPHC port samplers collect additional information on commercial fishing trips and catch composition. They are stationed in about a dozen ports in Washington, British Columbia, and Alaska that collectively account for the majority of landings from every regulatory area. For as many trips as possible, port samplers record the areas fished, amount of gear set and hauled, and catch by copying the skipper's logbook or interviewing the skipper. These records are combined with fish ticket data to calculate commercial catch per unit effort (CPUE) in each area.

Port samplers also obtain a carefully chosen random sample of (presently) about 1500 fish from each regulatory area, from which the length and age composition of the commercial landings can be estimated (Clark 2006a, Clark et al. 2000). From 1963 through 1990, in order to save money, the lengths of fish in the sample were not actually measured but predicted from a regression of body length on otolith size (Clark 1992a), which complicates the assessment in some ways. Since 1991 samplers have measured the lengths.

Setline survey data

Except for a hiatus in the years 1987-1992, IPHC has conducted systematic setline surveys since 1977, with both the frequency and coverage of surveys increasing over the years. Before 1996, no surveys were done in Areas 3B and 4. Since 1997, most areas have been surveyed in their entirety nearly every year. In recent years survey stations have been placed on a square 10 nautical mile (nmi) grid covering the entire continental shelf between 20 and 275 fathoms (fm). Between four and eight standard skates (100 baited hooks each) have been set at each station. Figure 3 shows the survey stations fished successfully in Area 3B in 2004. All halibut in the catch are measured, and a random sample (of target size 2000 per area) is collected for age, sex, and maturity determination.

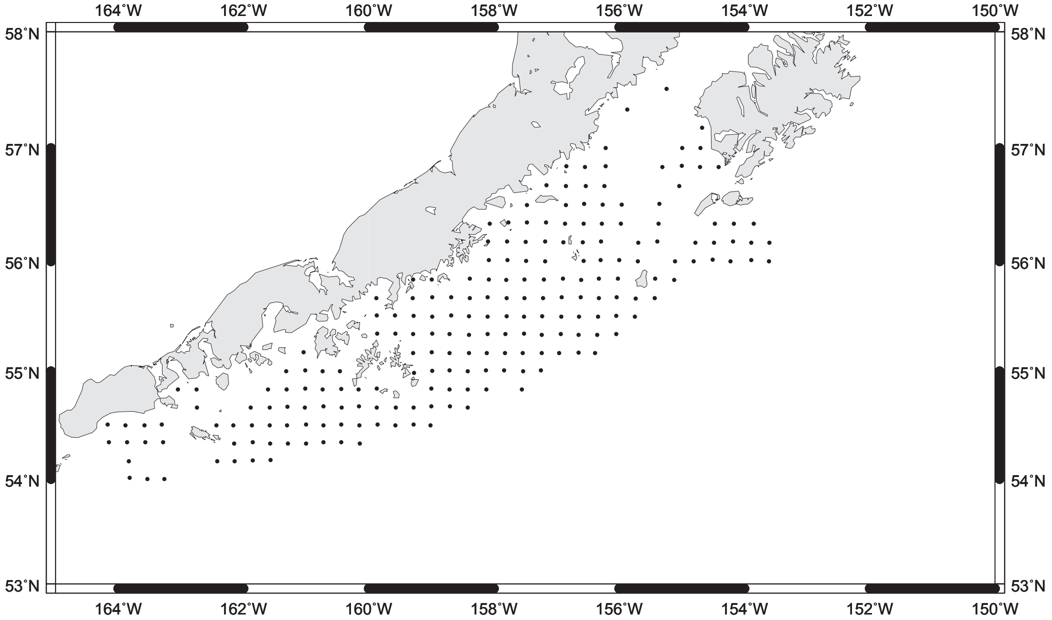


Figure 3. IPHC setline survey stations in Area 3B in 2004. Some sets in the eastern part were ineffective due to heavy shark damage.

Bycatch estimates

Halibut taken as bycatch in other groundfish fisheries must be returned to the sea, and a proportion of them die in the process. Both Canada and the United States place observers aboard fishing vessels to estimate the amount and length composition of the halibut bycatch, and to assess the condition of halibut before being discarded. These condition factors are used to predict mortality. The bycatch estimates available for the assessment are therefore estimates of bycatch mortality in number by length; no age data are collected.

Bycatch varies greatly among regulatory areas in both amount and size composition (Fig. 4). In Areas 3 and 4 where there are large trawl fisheries and large numbers of juveniles, the bycatch is large and has a modal length around 50-60 cm. In Area 2 where there is less trawling and fewer juveniles, the bycatch is much smaller and has a modal length of 70-80 cm. In 2004 bycatch mortality totaled 12 million (net) pounds (Williams 2005a), about evenly divided between fish larger and smaller than the commercial minimum size limit (81 cm).

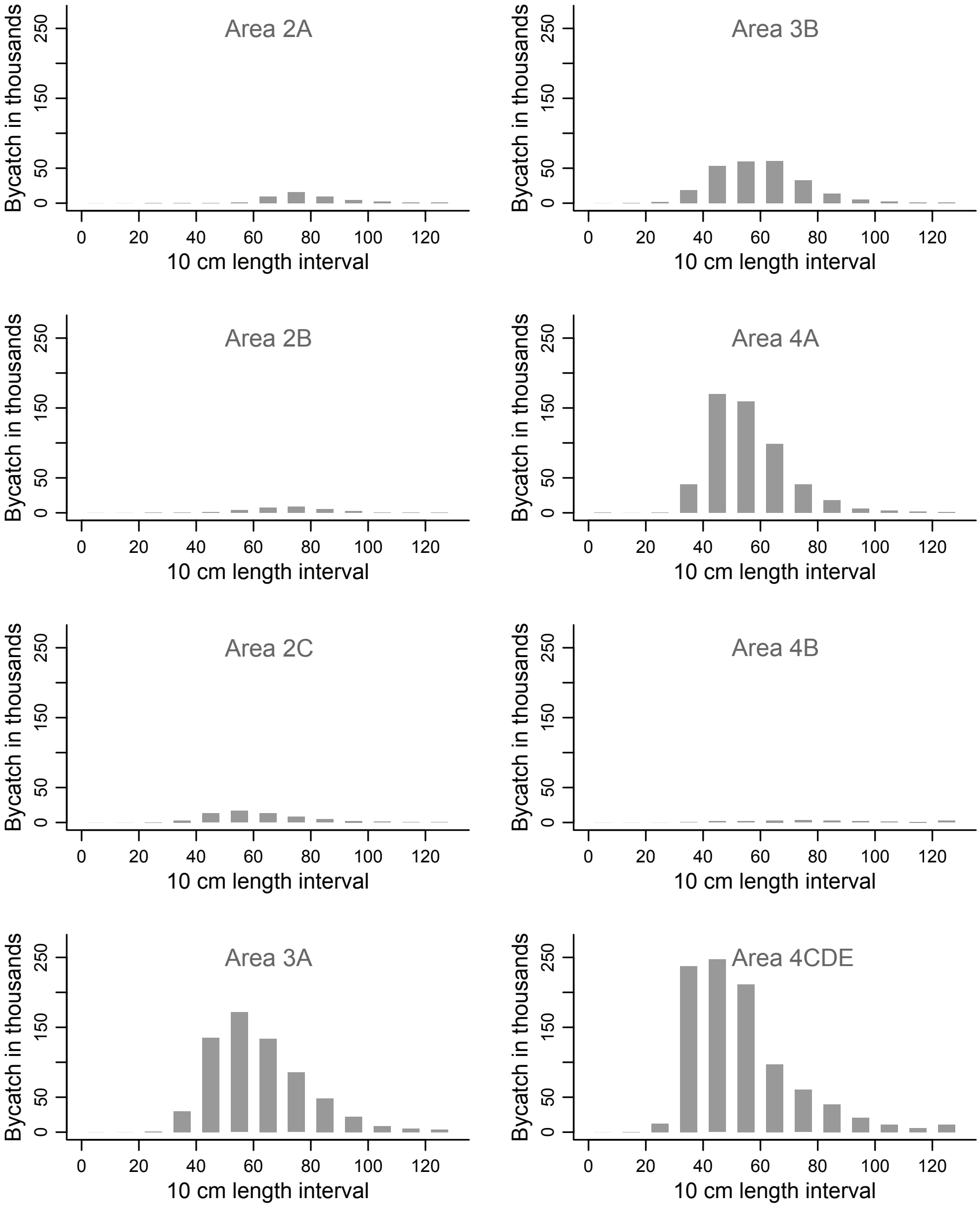


Figure 4. Bycatch in number by 10 cm length interval by regulatory area in 2004.

Sport catch estimates

There are substantial sport fisheries in Areas 2 and 3A. Sport catches in U.S. waters are estimated in various ways by the states of Oregon, Washington, and Alaska. The Canadian Department of Fisheries and Oceans (DFO) estimates the British Columbia sport catch. Length frequency data are available for most but not all jurisdictions; age samples only from Alaska. The length frequencies of sport catches are very similar to the length frequencies of IPHC setline survey catches (Fig. 5). In 2004 sport catches totaled nine million (net) pounds (Blood 2005). Fish below the 81 cm commercial size limit made up about 30% of the sport catch in number but only about 10% in weight.

Subsistence catch estimates

Both Canada and the United States authorize some fishing for subsistence or personal use apart from sport fishing. The catches in weight are reported but no length or age data are collected. Because these are all hook-and-line fisheries, they are assumed to have length frequencies similar to IPHC setline survey catches, like the sport catches. In 2004 subsistence catches totaled 1.4 million (net) pounds (Williams 2005b).

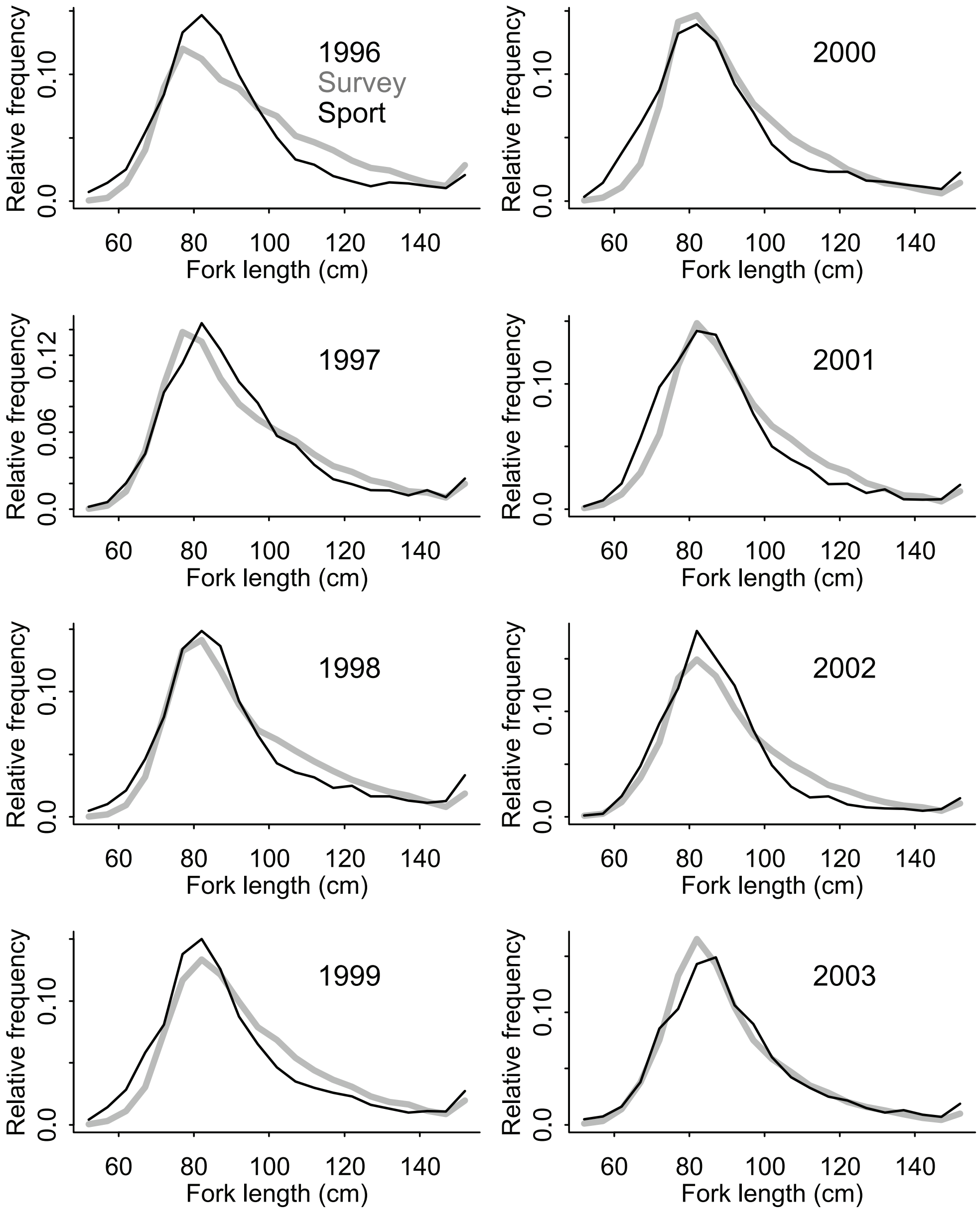


Figure 5. Length frequencies of sport catches and IPHC setline survey catches in Area 3A.

Data compilation and preprocessing

Commercial data selection

Commercial setline gear consists of a long, stout groundline with baited hooks on 1-4 ft gangions (leaders) attached at 5-20 ft intervals. The gangions can be permanently tied to beckets on the groundline, in which case the gear is called fixed-hook or just fixed; or they can be attached with snaps each time the groundline is set and removed when it is hauled back, in which case the gear is called snap-hook or just snap. In controlled experimental sets fixed and snap gear have equal catch rates (Myhre and Quinn 1984), but in the commercial fishery fixed gear generally has higher catch rates.

In almost all areas the assessment uses only fixed-hook commercial CPUE, for the following reasons:

(i) To avoid variations in commercial CPUE due to changing proportions of fixed and snap effort.

(ii) Because fixed-hook data are available back to the beginning of the fishery, whereas snap gear did not appear until the 1950s.

(iii) Because the spacing of hooks on snap gear can be quite variable, which means that the average hook spacing recorded in logbook data is imprecise. This complicates the hook spacing adjustment (explained below).

The exception to this rule is Area 2B, where at present the great bulk of the catch is taken with snap gear. Fortunately this is also the area where the relationship between fixed and snap CPUE is most consistent, with fixed CPUE being close to 135% of snap CPUE year after year (Clark 2002a). Both kinds of gear are used in the Area 2B assessment, with snap CPUE scaled up by 1.35.

Setline survey data selection

The early survey data (before 1993) have some features that require care in compiling a time series; these are explained in Appendix A. Generally all survey data are used in the assessment, except in Area 2B where only the area north of Cape Scott (containing about 90% of the Area 2B abundance) is consistently surveyed. Where extensions of the survey within areas have changed average CPUE (Clark 2002b), that effect is estimated within the assessment model (as a change in survey catchability) rather than being estimated externally and used to adjust the raw data.

Hook type and hook spacing adjustments

The commercial fishery switched from J-hooks to much more effective circle hooks (C-hooks) in 1983; the survey followed suit in 1984, when both hook types were fished. The 1984 survey data showed that C-hook catchability was more than twice that of J-hooks and that length-specific selectivity was also different (Sullivan et al. 1999, Appendix 1). In the IPHC database all of the J-hook effort—commercial and survey—is divided by 2.2 to make the J-hook catch rates comparable with C-hook catch rates. That adjustment is removed when the assessment data are compiled so that the assessment uses raw J-hook data and estimates the hook change effect separately for commercial and survey CPUE in each regulatory area. The reason for doing that is to allow for differences among areas and between commercial and survey CPUE in the effect of the hook change. The 1983 commercial data contain an unknown mixture of hook types and are not used in the assessment.

Setline CPUE is also affected by hook spacing, with catch per hook generally increasing with spacing (distance between hooks). Hamley and Skud (1978) conducted experiments using J-hook gear and spacings in the range 10-40 ft, and found that catch per hook at a spacing of H ft was $1.52 \cdot \left(1 - \exp\left(-0.06 \cdot H\right)\right)$ times catch per hook at a spacing of 18 ft (Fig. 6). Recent

Figure 6. The hook spacing adjustment applied to survey and commercial CPUE.

analysis (Clark 2006b) has shown that this formula describes the effect of hook spacing in the present-day commercial fishery quite well. This adjustment is incorporated in the data stored in the IPHC database and used in the assessment without change. The adjusted catch per hook is also multiplied by 100 so that the effort data in the database all refer to a 100-hook skate with 18 ft spacing, called an effective skate.

Estimation of the catch at age

The weights of fish in the commercial sample in each area/year are estimated from their lengths using a length-weight relationship (Clark 1992b). The mean weight of fish in the catch is estimated as the mean weight in the sample, and the number of fish in the catch as the reported weight of landings divided by the mean weight. The catch at age is estimated by applying the age composition of the sample to the total catch in number.

Estimation of the sex composition of commercial landings

Because females are larger than males and commercial selectivity is determined by length, females and males are distinguished in the assessment model. The sex composition of the survey catches is observed and recorded, but fish in the commercial catch are eviscerated at sea, so port samplers cannot determine sex when they sample the landings. For the assessment, the sex composition of the landings at each age, and the size composition of females and males at each age in the landings, are estimated from survey catches in the same area and same year (or close to it). Details are given in Appendix B. This procedure limits the assessment in each area to those years with survey data. In Areas 2 and 3A the assessment can be extended back to 1974; in Areas 3B and 4 only to 1996.

Estimation of age misclassification

Ages of sampled fish are determined by reading their otoliths (Forsberg 2001). Before 2002 surface readings were done, and these tended to underestimate the ages of fish older than 12 years. Since then all otoliths have been broken and burned, which gives more accurate readings.

In addition to being biased, surface readings have a large variance. The assessment model incorporates a misclassification matrix to predict the observed surface age compositions from the calculated true age compositions. This step is called "smearing" the true age compositions because the effect of age reading error is to redistribute some of the fish of each age group to neighboring ages. A second misclassification matrix is used to account for the effect of the much smaller variance of break-and-burn readings on observed age compositions beginning with 2002. Details are given in Appendix C.

Estimation of size at age

In the assessment model, setline selectivity is treated as a function of observed mean size at age by sex in setline survey catches. This is not the true mean size at age in the stock because the gear selects for larger fish, but it can be used as an alternative size metric to predict selectivity at age. Survey data are used because the sexes are distinguished and the catches include fish below the minimum commercial size limit of 81 cm. Size at age has decreased greatly since the mid-1980s (Fig. 7).

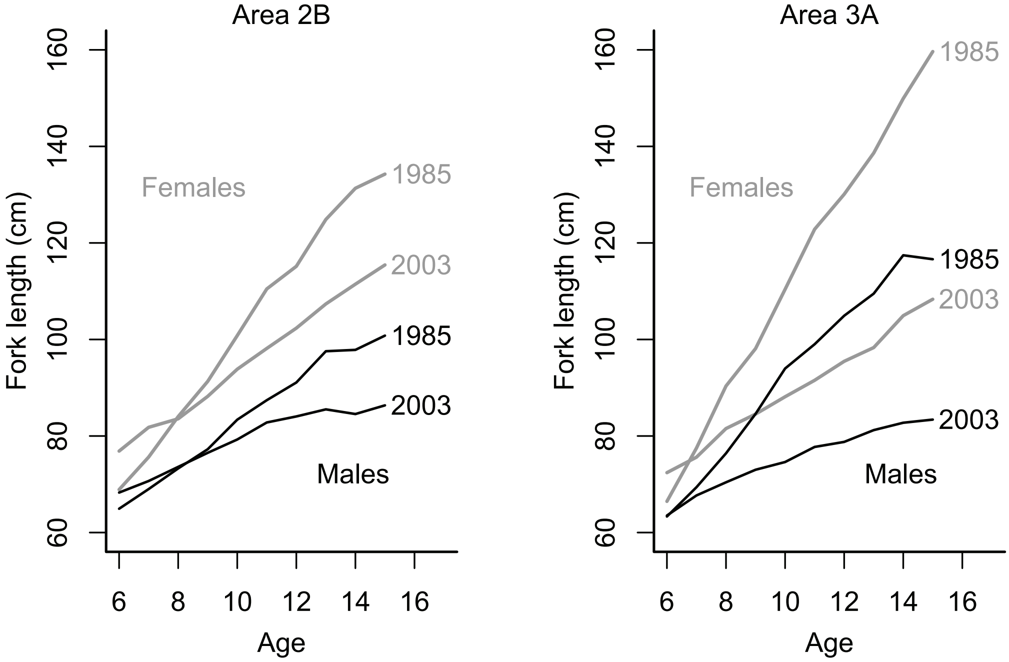


Figure 7. Mean length at age in IPHC setline surveys, by sex, in 1985 and 2003.

Survey data are used to estimate mean weight at age/sex in the survey and in the sport and subsistence fisheries (needed for fitting to reported catches in weight). Mean weight at age/sex in the commercial landings is estimated from the commercial data, broken down by sex as described above.

Estimates of the mean and standard deviation of size at age by sex in the bycatch are also needed for fitting the observed bycatch at length, which is mostly taken in trawl fisheries. Fish younger than age 6 are rare in setline catches, but trawl fisheries catch many younger fish. Among age groups that appear in both setline and trawl catches, mean size at age is lower in trawl than setline catches. The values of size at age/sex in the bycatch are based on NMFS trawl survey data. For ages 1 though 6, the values are the same for females and males and they are:

	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6
Mean length (cm)	15	25	35	45	52	59
Standard deviation (cm)	3.0	4.0	4.5	5.0	5.5	6.0
Mean weight (net lb)	0.05	0.20	0.80	1.70	2.65	3.90

The observed trawl survey values are used in the assessment for ages 1-5. For ages 6+, the mean and standard deviation of length at age/sex in the bycatch are 90% of the age/sex-specific setline survey values, and mean weight is 80% thereof. These multipliers are based on regressions of trawl values on setline survey values in areas and years that had both kinds of survey data. They allow the distribution of size at age in the bycatch to be estimated from setline survey data in areas where there are no trawl survey data.

Because of the bias in surface readings, size at age data are inaccurate for older fish through 2001. In order to have a consistent series for predicting selectivity in the assessment, an adjusted set of size at age estimates is calculated that attempts to correct for the underestimates of age in surface readings. It is not entirely successful in that even the adjusted series show some discontinuities in size at age between 2001 and 2002, but mainly among older and larger fish that are mostly or fully selected anyway.

Estimation of sampling variances

A non-zero sampling variance is estimated for every data point to which the assessment model is fitted, the main types being commercial catch at age/sex, commercial CPUE, survey age/sex composition, and survey CPUE. The rules used are:

(i) A multinomial variance is calculated for proportions at age: $\hat{V}\left(\hat{p}\right)=\hat{p}\cdot\left(1-\hat{p}\right)/n$, where n is the sample size for that area/year.

(ii) If a sample proportion is zero, a variance is calculated with $\hat{p}=0.005$.

(iii) The setline survey stations fished in a given area/year are treated as a simple random sample, and the mean and standard deviation of catch per skate are calculated with the standard formulas. This must overestimate the variance to some extent because the stations are placed systematically.

(iv) Commercial CPUE in a given area/year is assigned a coefficient of variation (CV) of 0.05, based on the scatter of year-to-year values about a data smoother. The amount of logbook data is normally so large that sampling variance in the strict sense is practically nil.

(v) The estimated proportion female in a given area/year/age stratum (Appendix B) is also assigned a CV of 0.05, also based on the scatter of data points about the fitted logistic curves.

(vi) The estimated number of fish in a given 5 cm length interval in the bycatch $\hat{B}_l$ is assigned a CV of $10\Big/\sqrt{\hat{B}_l}$, on the grounds that the unknown underlying sample sizes are on the order of 1% of the bycatch.

Assessment methods

IPHC has set catch limits on the basis of quantitative stock assessments since 1932. In the early years the rules came from the pioneering work of Thompson and Bell (1934). Since 1982 the annual assessment has consisted of fitting an age-structured model to commercial (and later survey) data, a procedure developed at the Commission (Deriso et al.1985) and later adopted by most agencies in North America. Clark (2003) provides a history of the staff's modeling work.

At present a separate model fit is done for each regulatory area except Area 2A and Area 4CDE (Fig. 1), which are handled differently as explained at the end of this section. The area-specific model fits assume that the fish in each regulatory constitute a closed population; i.e., that there is no immigration or emigration among fish that have reached catchable size (60-80 cm). Historical marking data support this view (Trumble et al. 1990), but some doubts have been raised by recoveries of passive integrated transponder (PIT) tags from an experiment that is in progress at time of writing (Clark 2006c). If it turns out that there is significant net migration among areas, it may become necessary to enlarge the scope of the assessment, but that would not entail a change in the methods or model.

Evolution of present assessment methods

From 1982 through 1994, the halibut stock assessment relied on CAGEAN, a simple age-structured model fitted to commercial catch-at-age and catch-per-effort data. The constant age-specific commercial selectivities used in the model were fundamental model parameters, estimated directly.

Beginning in the late 1980s, halibut growth rates in Alaska declined dramatically. As a result, age-specific selectivity decreased. CAGEAN did not allow for that, and by the mid-1990s was seriously underestimating abundance. In effect, it interpreted lower catches as an indication of lower abundance, whereas the real cause was lower selectivity. Incoming year classes were initially estimated to be small, but in subsequent years' assessments those estimates would increase when unexpectedly large numbers of fish from those year classes appeared in the catches. The year-to-year changes in the stock trajectory shown by the assessment therefore developed a strong retrospective pattern. Each year's fit showed a steep decline toward the end, but each year the whole trajectory shifted upward.

The staff sought to remedy that problem by making selectivity a function of length in a successor model developed in 1995. It accounted not only for the age structure of the population, but also for the size distribution of each age group and the variations in growth schedule that had been observed. The fundamental selectivity parameters in this model were the two parameters of a function (the left limb of a normal density) by which the selectivity of an individual fish was determined from its length. The age-specific selectivity of an entire age group was calculated by integrating length-specific selectivity over the estimated length distribution of the age group, and that age-specific selectivity was used to calculate predicted catches. The new model was fitted to both commercial data and IPHC setline survey data, with separate length-specific selectivity functions. Commercial catchability and selectivity were allowed to drift slowly over time, while survey catchability and selectivity were held constant (Sullivan et al. 1999).

When this model was fitted to data from Area 2B and Area 3A, quite different length-specific selectivities were estimated, which suggested that fishery selectivity was not wholly determined by the properties of the gear and the size of the fish but also depended on fish behavior (e.g., migration). These behavioral elements are likely to be more related to age than size. The age of sexual maturity, for example, remained virtually the same in Alaska despite the tremendous decrease in growth, so the size at maturity is now much smaller than it was. While size must affect selectivity, it was thought that age was also influential.

To allow for that, the model was fitted in two ways. The original form was called the "length-specific" fit, because a single set of estimates of the two parameters of the length-based survey selectivity function was used in all years. In a second form, called the "age-specific" fit, the parameters were allowed to drift over time (like the commercial selectivity parameters), but they were required (by a heavy penalty) to vary in such a way that the integrated age-specific selectivities calculated in each year remained constant over time.

The usual diagnostics gave little reason to prefer one fit over the other. Goodness of fit was similar: good for both in 2B, not so good for either in 3A. The retrospective behavior of both fits was dramatically better than that of CAGEAN and quite satisfactory in all cases, although the length-specific fit was more consistent from year to year in 3A and the age-specific fit was more consistent in 2B (Clark and Parma 1999). The two fits produced very similar estimates of abundance in Areas 2B and 2C, but in 3A the length-specific estimates were substantially higher, so out of caution the staff catch limit recommendations were based on the age-specific fit through 1999.

The assessment model was simplified and recoded as a purely age-structured model in 2000 to eliminate some problems associated with the modeling of growth and the distribution of length at age. It retained the option of modeling survey selectivity as a function of mean length at age (observed not predicted), but the production fits continued to be based on constant age-specific survey selectivity, estimated directly as a vector of age-specific values rather than as a parametric function of age.

The fit of this model to Area 3A data in 2002 showed a dramatic retrospective pattern, similar to the pattern of successive CAGEAN fits in the mid-1990s. Treating setline survey selectivity as length-specific rather than age-specific largely eliminated the pattern, just as freeing up age-specific commercial selectivity had improved the retrospective behavior of CAGEAN. Accumulated data showing very similar trends in CPUE at length in IHPC setline surveys and NMFS trawl surveys provided further evidence that setline selectivity is, after all, determined mainly by size rather than by age.

Another anomaly of the 3A model fit in 2002 was the unexpectedly large number of old fish (age 20+) in the last few years' catches. This was found to be the result of an increase in the proportion of otoliths read by the break-and-burn rather than surface method. Surface readings tend to understate the age of older fish, and IPHC age readers had been gradually doing more and more break-and-burn readings as the number of older fish in the catches increased. The poor model fit at these ages indicated a need to deal explicitly with the bias and variance of both kinds of age readings.

An entirely new model was written for the 2003 assessment (Clark and Hare 2004). Both commercial and survey selectivity were parameterized as piecewise linear functions of mean length at age in survey catches, and were required to reach an asymptote of one at or before a length of 130 cm. Because females are larger than males, all of the population accounting and predictions were done separately for each sex. (The age/sex/size composition of the commercial landings was estimated external to the assessment for this purpose.) The observed age compositions (surface or break-and-burn) were predicted by applying estimated misclassification matrices to the age distributions. Even in its most parsimonious form—with just one survey and one commercial selectivity schedule for both sexes in all years—this model achieved very good fits to the sex-specific observations and good retrospective performance. It also produced somewhat higher estimates of average recruitment and recruitment variability. With this simple model it was feasible do standalone analytical assessments of abundance in Areas 3B, 4A, and 4B for the first time, using data from 1996-2003.

Only two minor changes were made for the 2004 assessment, and neither had a significant effect on the estimates of abundance. First, both the 2004 PIT tag recoveries (Clark and Chen 2005) and a reanalysis of earlier wire tag data (Clark 2005) indicated that commercial selectivity is not always asymptotic; it appeared to be more dome-shaped in Area 2B and more ramp-shaped in Area 3A. Fitting the assessment model with free-form selectivity schedules showed much the same thing for commercial selectivity, namely an assortment of shapes beyond 120 cm. Nevertheless a schedule that reaches an asymptote of one at 120 cm is a good approximation to and compromise among the free estimates, and using an asymptotic commercial schedule is desirable for computing exploitable biomass and reporting harvest rates, so that is what was used in the assessment. All of the freely estimated survey selectivities either level out or increase after 120 cm. Freely estimated survey selectivities present no practical difficulties, so they were estimated that way in the assessment, and most of the estimates were ramp-shaped.

Treatment of natural mortality

An age-structured assessment model estimates the size of each year-class by back-calculating the sum of all removals from it, including commercial catches, other directed catches, bycatch, and natural deaths. The natural deaths can be a large fraction of the total removals, so the natural mortality rate used in the calculations has a large effect on estimates of absolute abundance.

Natural mortality rates are notoriously difficult to estimate directly. Early halibut catch curve analyses gave estimates of the instantaneous natural mortality rate (M) ranging from 0.15 to 0.25 with an average around 0.20 (IPHC Staff 1960). Myhre (1967) estimated total mortality from tagging experiments and then obtained area-specific estimates of M by regressing total mortality on fishing effort. His estimates were around 0.30, which he considered unrealistically high. All of these estimates are questionable because they rely on equilibrium assumptions that probably did not hold. Chapman et al. (1962) stated that any value in the range 0.15-0.20 would be consistent with the available data, and Myhre (1974) accordingly opted for a value of 0.175 when doing the calculations to locate an optimum minimum size limit.

An age-structured assessment model will usually provide equally good fits for a wide range of natural mortality rates, so in practice the analyst has to pick a value outside the assessment and use that in the model calculations for better or worse. That was (and still is) the case for the halibut assessment, and during the 1980s and most of the 1990s the staff resolutely used the value 0.20.

Clark (1999) analyzed the effects of an erroneous estimate of natural mortality on yield recommendations when the same erroneous estimate is used both in the stock assessment where absolute abundance is estimated and in the dynamic fishery simulations where an optimum harvest rate is chosen. He found that the yield recommendations were quite robust to error in the estimate of natural mortality, but that the cautious policy was to select a value toward the low end of the plausible range. The working value in the assessment was therefore lowered to 0.15 in 1998, and there it has remained.

Recently Lester et al. (2004) formulated a simple energetic model of growth and reproduction and from that derived equations for the age at maturity T and gonadosomatic index g (gonad weight as a proportion of somatic weight) that maximize fitness given the rate of natural mortality M and the age intercept t_1 of a linear function $L_t = h_1 \cdot (t - t_1)$ describing length at age L_t during the years before the age of maturity. These equations can be inverted to estimate the natural mortality rate of halibut from the observable parameters T, g, and t_1. Unlike other predictors of natural mortality based on life history parameters, this procedure does not depend in any way on empirical relationships between life history parameters and published estimates of natural mortality in various stocks that are themselves highly questionable. For halibut, a straight line through the origin represents juvenile growth quite well, so $t_1 \approx 0$. The age at maturity of females has remained at eleven years despite large changes in size at age (Hare and Clark 2005). The gonadosomatic index of halibut is about 0.18 (Schmitt and Skud 1978). Lester et al. (2004) found that $(T - t_1) \approx 1.95 / (e^M - 1)$, implying $M = 0.163$ for halibut. They also found that $g \approx 1.18 \cdot (1 - e^{-M})$, implying $M = 0.165$. These estimates are purely theoretical and not at all precise, but they do provide some reason for believing that the working value $M = 0.15$, while uncertain, is not so far the true value that the yield recommendations are wrong.

Present model structure

The assessment model is a conventional age-structured model. The parameters estimated are numbers at age in the first year (1974 in Areas 2 and 3A, 1996 elsewhere), subsequent recruitments, survey catchability and selectivity, commercial catchability, and selectivity and annual fishing mortality rates for all fisheries (commercial, sport, subsistence, and bycatch). The model calculates abundance and catches for ages 1-30+, with age 30+ being the plus group where all older fish are accumulated. It predicts commercial catch at age, total commercial CPUE in number and weight, survey age composition, total survey CPUE in number and weight, bycatch in number at length, and sport and subsistence catch in weight. These predictions refer to true age compositions. They are multiplied by an age misclassification matrix (Appendix C) to predict the observed age compositions, which are influenced by bias and variance in the age readings. The predictions are calculated for ages 1-20+ for years with surface ages (through 2001) and for ages 1-25+ for recent years with break-and-burn ages because that is how the data are tabulated. Carrying the model calculations out to age 30+ assures that the effect of age misclassification is properly accounted for in predicting the observed numbers in and near the plus group in the data.

Modeling of catchability

Catchability coefficients are estimated for the commercial fishery and the setline survey for the purpose of predicting the respective CPUE series. Changes in catchability over time are allowed by a flexible system of waypoints and paths that can provide a piecewise linear approximation of any trajectory. For each CPUE series, a catchability coefficient can be estimated for each of a specified set of years—the waypoints. In an Area 3A assessment, for example, separate commercial catchabilities could be estimated for 1974 (the first year in the data, when the fishery used J-hooks exclusively), 1984 (the first year of pure C-hook fishing), and every four years thereafter (1988, 1992, …). During the years between waypoints, the path of the catchability coefficient can be either flat or interpolated. If flat, the catchability stays at the value of the initial waypoint until the next waypoint and then changes abruptly. This would be appropriate for the 1974-1982 period of J-hook fishing before the abrupt change in hook type in 1983. If the path is interpolated, the catchability is interpolated between the waypoint values. This would be appropriate to describe the gradual, continuous changes in commercial selectivity beginning in 1984 as a result of technological changes, derby fishing, individual quotas, etc. The chosen waypoints and path types determine a catchability coefficient for every year in the data series. The actual estimates (Fig. 8) show the large increase in catchability that resulted from the change to C-hooks in 1984, the gradual decline in catchability during the derby fishery of the later 1980s and early 1990s, and the increase in catchability after the adoption of individual quotas in 1995. Because the survey follows a standard protocol, survey catchability is assumed to be constant apart from the effect of changing to C-hooks in 1984.

Modeling of selectivity

The setline selectivity of an age/sex group in a given fishery in a given area/year is modeled as a function of its observed mean length in setline survey catches (or trawl catches for bycatch). Instead of a parametric function, the model fits a piecewise linear approximation by estimating selectivity at 10 cm intervals from 60 to 130 cm and interpolating between them (Fig. 9). The value at 60 cm is used below there, and the value at 130 cm is used above there. This sort of function can assume a variety of forms, and it can be constructed so as to have a prescribed form. The options available in the model are:

(i) Asymptotic, with selectivity simply set to one at and after a specified length (i.e., selectivity is fixed not estimated at those lengths). Commercial setline selectivity is modeled this way.

(ii) Domed, with selectivity set to one at a specified length and required to decrease monotonically on either side of it.

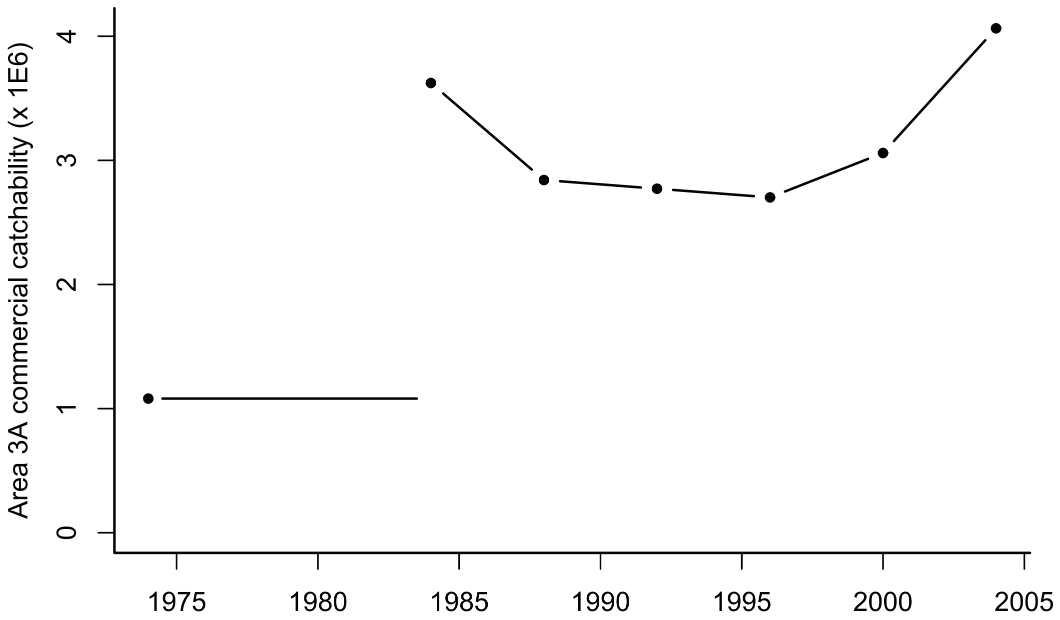


Figure 8. Estimated path of commercial catchability in Area 3A. A constant value is estimated for the J-hook period (1974-1982) and gradual changes from 1984 on, with catchability estimated at waypoints every 4 years and interpolated between.

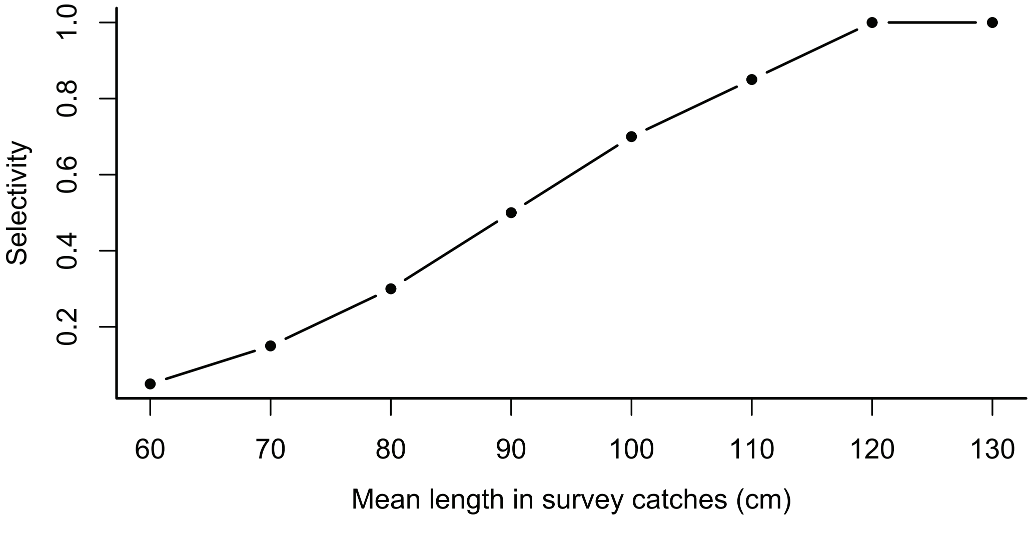


Figure 9. Setline selectivity at age/sex is modeled as a piecewise linear function of observed mean length in survey catches. The parameters estimated are the selectivities at the points spaced every 10 cm. Selectivity at intervening lengths is interpolated.

(iii) Free-form, with selectivity set to one at a specified length and allowed to have any value (including values greater than one) elsewhere. Setline survey selectivity is modeled this way, with selectivity set to one at 120 cm. So is bycatch selectivity, with selectivity set to one at 60 cm and estimated elsewhere every 10 cm from 0 cm to 120 cm.

Changes in selectivity parameters over time are modeled by the same system of waypoints and paths described above for catchability. Commercial selectivity is lower than survey selectivity for young fish because of the effect of the commercial minimum size limit of 81 cm, but it is not necessarily zero for age/sex groups whose mean length is less than 81 cm because some fraction will be legal-sized. The sport and subsistence fisheries are assumed to have the same selectivity as the setline survey.

Catch equations and survivorship

All sources of mortality are modeled as competing exponential rates, and predicted catches are calculated with the Baranov catch equation. Specifically, let the subscript c denote the commercial fishery, r the sport fishery, p the subsistence fishery, and b the bycatch fisheries. The selectivities of a given age/sex group g (Sel_{cg}, Sel_{rg} etc.) in a given area/year are determined by its mean length in setline (or trawl) survey catches as explained above. Instantaneous fishing mortality rates are then just the product of the selectivity and the corresponding full-recruitment fishing mortality in that area/year (F_c, F_r etc.), or $F_{cg} = Sel_{cg} \cdot F_c$ and so on. The total instantaneous mortality rate is $Z_g = F_{cg} + F_{rg} + F_{pg} + F_{bg} + M$. If the number of survivors at the beginning of the year is N_g, the average number during the year is $\bar{N}_g = N_g \cdot \left(1 - \exp\left(-Z_g\right)\right) / Z_g$ and the predicted catches in number are $C_{cg} = F_{cg} \cdot \bar{N}_g$ and so on. The number of survivors at the beginning of the next year is $N'_g = N_g \cdot \exp\left(-Z_g\right)$. If commercial catchability in that area/year is q_c, predicted commercial CPUE in number for that group is $CPUE_g = q_c \cdot Sel_{cg} \cdot \bar{N}_g$ and predicted commercial CPUE in weight is just $CPUE_g \cdot \bar{w}_{cg}$ where $\bar{w}_{cg}$ is observed mean weight for that group in the commercial landings. Predicted setline survey CPUE in total number and in weight of legal-sized fish are calculated the same way.

Using the Baranov catch equation (rather than treating the commercial fishery as a point removal as in some earlier models) permits the model code to be highly modular. Most of it consists of fishery-specific routines that determine catchabilities, selectivities, and fishing mortality rates from the specified waypoints and paths. Once that has been done, the core catch and survivorship calculations take only a few lines, so the really critical section of the code is very robust.

Summary of parameters estimated and data predicted

The table below lists the model parameters estimated along with the data predicted for each area. The subscript y denotes year, a age, s sex, f fishery type, and k length interval.

Parameter	**Description**	**Data predicted**
$InitN_{as}, a = 2,...,30$	Number at age/sex in the first year.	All.
R_y	Recruitment at age 1 in each year (except for the last six or so).	All.

Parameter	Description	Data predicted
F_{fy}	Full-recruitment fishing mortality for each fishery in each year.	Catch at age/sex in all fisheries (commercial, sport, subsistence, bycatch).
$CSel_{ky}, k = 60, 70, \ldots 120;$ $y = \text{(e.g.) } 1974, 1984, 1988, \ldots$	Commercial selectivity at 10-cm waypoints in chosen waypoint years.	Commercial catch at age/sex, total commercial CPUE in number and weight.
$CQ_y, y = \text{(e.g.) } 1974, 1982, 1984, 1988, \ldots$	Commercial catchability at chosen waypoint years.	Total commercial CPUE in number and weight.
$SSel_{ky}, k = 60, 70, \ldots, 130;$ $y = \text{(e.g.) } 1974, 1984$	Survey selectivity at 10-cm waypoints in chosen waypoint years, normally one schedule for J-hooks and one for C-hooks.	Survey age/sex composition, total survey CPUE in number and weight, sport and subsistence catch at age/sex.
$SQ_y, y = \text{(e.g.) } 1974, 1984$	Survey catchability, normally one for all J-hook years and one for all C-hook years.	Total survey CPUE in number and weight.
$BSel_{ky}, k = 0, 10, 20, \ldots 120;$ $y = \text{(e.g.) } 1974,\ 1979,\ 1984, \ldots$	Bycatch selectivity at 10-cm waypoints in chosen waypoint years.	Bycatch at age/sex.

Error structure and log likelihood

The error structure and likelihood are based on Fournier et al. (1990). In summary, the model is fitted by minimizing the sum of scaled, squared deviations between the model predictions and the observations, meaning that all the observations are treated as independent normal random variables. In addition, the squared deviations are calculated in a way that makes the estimates robust to the few outliers by reducing their influence.

Variance scalers

Let Y_i denote an observation in the data, which might be a commercial catch in number at age/sex or a proportion at age/sex in the survey or an annual commercial CPUE, and let s_i^2 denote the estimated sampling variance of Y_i, and $\hat{Y}_i$ the model prediction of it. The usual procedure is to weight each squared deviation by the inverse of its variance, or:

$$D_i^2 = \frac{\left(\hat{Y}_i - Y_i\right)^2}{s_i^2}$$

If the model were correctly specified and the observations contained only sampling error, the $\langle D_i \rangle$ would be standard normal random variables, so the mean error of the predictions would be zero and the root mean squared error would be one. This is far from true; the mean error is near zero but the root mean squared errors are 2 to 3 depending on the data type (Table 1). What this means is that sampling variance accounts for a quarter or less of the total error variance.

To stabilize and standardize the variances of the deviations, each of the sampling variances of the observations is multiplied by a variance scaler τ_i^2 and a scaled squared deviation d_i^2 is computed with that weight:

$$d_i^2 = \frac{\left(\hat{Y}_i - Y_i\right)^2}{\tau_i^2 \cdot s_i^2}$$

The $\langle \tau_i \rangle$ can be modeled in various ways, and they can be estimated internally as parameters (Fournier et al. 1990) or set externally (Maunder and Watters 2003). In the halibut assessment the working values of the $\langle \tau_i \rangle$ are set to the average values of the root mean squared errors for each data type in unscaled fits; they are shown in the last column of Table 1.

Apart from outliers, the scaled deviations $d_i = \left(\hat{Y}_i - Y_i\right) / \left(\tau \cdot s_i\right)$ should have approximately a standard normal distribution, and they do. In Area 3A (Fig. 10b), the commercial and survey catch at age deviations match a standard normal distribution quite well, and there is only one deviation greater than 5. Outliers are more numerous among the Area 2B commercial catch at age deviations (Fig. 10a), with 17 (out of 475) deviations greater than 5 or less than -5. In both areas the commercial and survey CPUE deviations depart from a standard normal distribution, but this is mainly the result of the small sample sizes (only 25-30 points).

Table 1. Root mean squared errors of various model predictions in unweighted fits using unscaled sampling variances. The working value of τ in the last column is used to scale the sampling variances when fitting the model normally.

Data type	Area 2B	Area 2C	Area 3A	Central value and working value of τ
Catch at age				
Total	2.6	2.3	2.4	2.5
Female	1.9	2.1	2.0	2.0
Male	1.8	2.3	1.7	2.0
Commercial CPUE				
Number	2.2	2.4	2.1	2.2
Weight	2.2	2.9	2.3	2.5
Survey proportion at age				
Total	2.6	2.7	2.6	2.5
Female	2.6	3.1	2.6	2.5
Male	2.9	2.3	2.3	2.5
Survey CPUE				
Total number	2.7	2.4	4.3	3.0
Legal-sized weight	2.8	2.4	5.3	3.0
Bycatch at length	2.6	2.0	4.3	3.0

It is somewhat inconsistent to treat the catch at age data as having a normal distribution when the sampling distribution is multinomial and the sampling variance is estimated using the multinomial formula. Fournier et al. (1990) give some practical and statistical reasons for doing so as a general practice. In the case of the halibut assessment the compelling reason is that the scaled deviations really do follow a normal distribution. This is not surprising. The individual sample proportions are binomial, and the normal distribution approximates the binomial well. Moreover, when the sampling variance accounts for only a small fraction of the total variance, one should not expect the shape of the distribution to be determined by the shape of the sampling distribution. The variance of observed proportions about the model predictions doubtless results from a combination of various kinds of process error and model misspecification, the sum of which could be expected to produce a normal distribution of deviations.

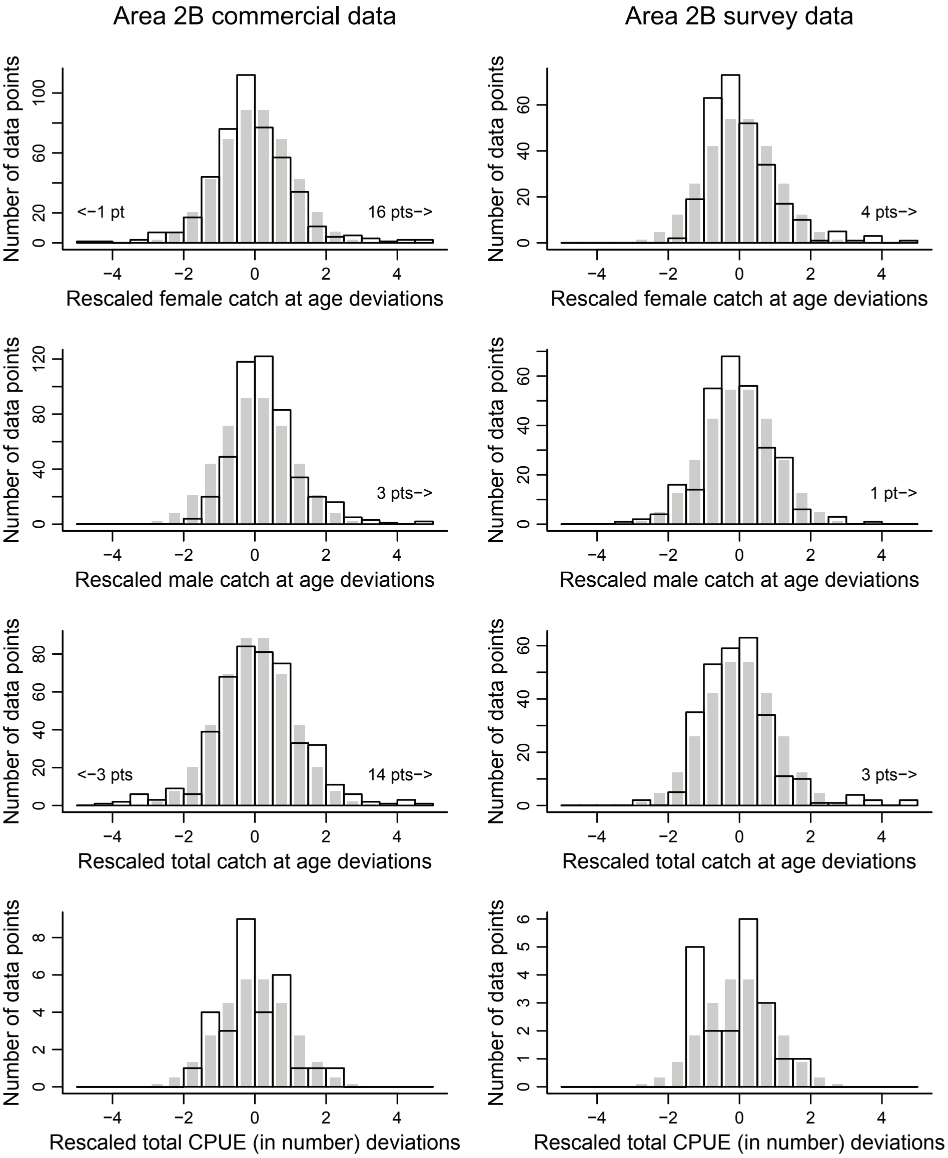


Figure 10a. Distribution of scaled deviations of model predictions from observations in the 2004 assessment of Area 2B. The white bars outlined in black show the actual distribution of deviations; the gray bars show a standard normal distribution for comparison. The notation “14 pts →” means there were 14 deviations larger than 5 that are not plotted.

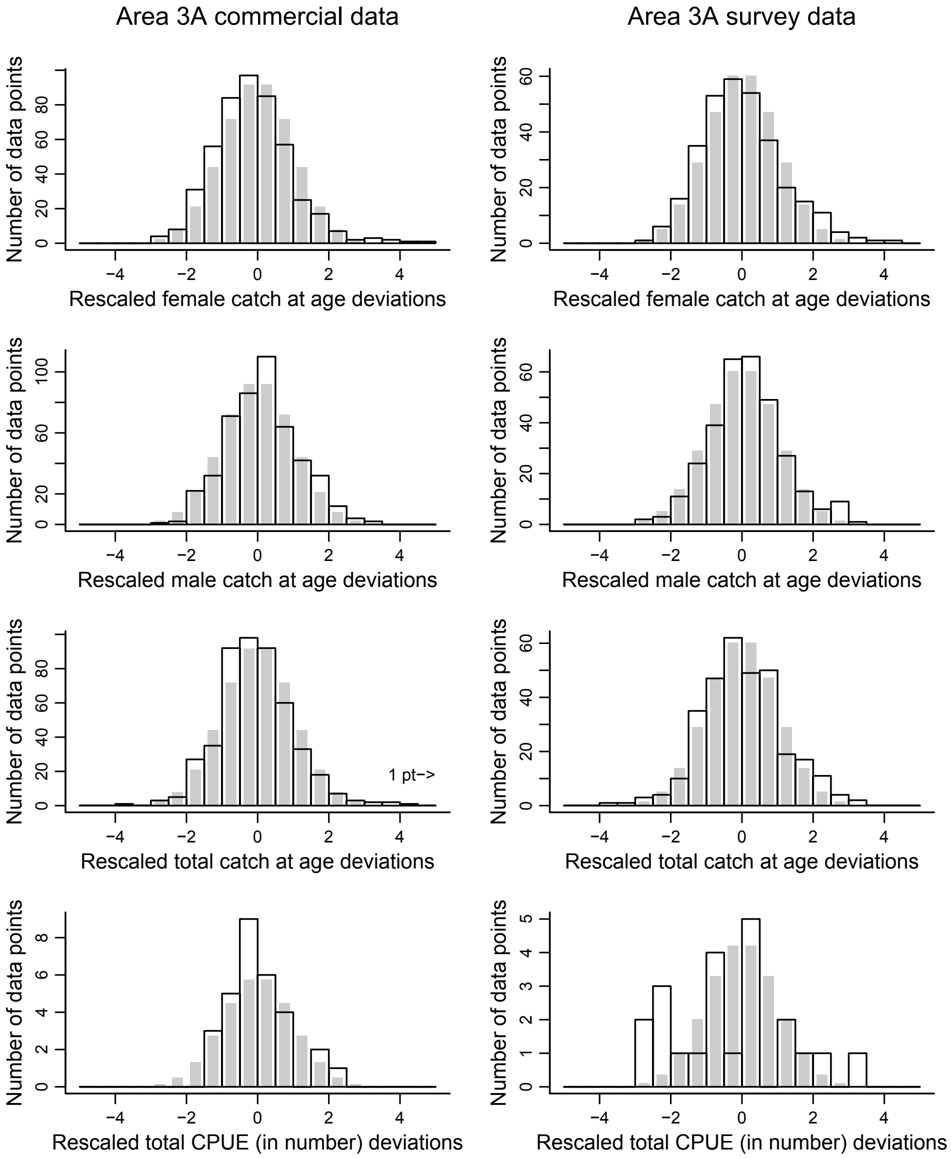


Figure 10b. Distribution of scaled deviations of model predictions from observations in the 2004 assessment of Area 3A. The white bars outlined in black show the actual distribution of deviations; the gray bars show a standard normal distribution for comparison. The notation "1 pt→" means there was 1 deviation large than 5 that is not plotted.

Robust estimation

While not numerous, the very large values among the scaled deviations when squared can make a large contribution to the sum of squares and have an inordinate influence on the estimates. To avoid giving that much weight to points that may just be outliers, the scaled deviations are run through a smooth function (Fig. 11) that is the identity function up to $|d| = 2.5$ and thereafter increases at a decreasing rate toward an asymptote at $|d| = 3$:

$$\text{for } |d| \le 2.5: \quad f(d) = d$$

$$\text{for } |d| > 2.5: \quad f(d) = \operatorname{sign}(d) \cdot \left(2.5 + 0.5 \cdot \left(1 - \exp\left(-\left(|d| - 2.5\right)/0.5\right)\right)\right)$$

The first derivatives of both portions of this function are $f'(d) = 1$ at $d = 2.5$, so it has a continuous first derivative, which is important for numerical minimization.

Application of this function has the effect of mapping all deviations greater than 2.5 onto the interval (2.5, 3), so they are still treated as improbably large deviations but not as impossibly large deviations. Fournier et al. (1990) accomplish the same thing by adding a small constant to the normal density function.

To obtain the robust estimates, the model is first fitted with the raw scaled deviations, including outliers, from an arbitrary starting point. (If robustified deviations were calculated at this point, all of the data would look like outliers.) The model is then refitted with robustified deviations using the first fit as the starting point. The parameter estimates are almost the same, but the extra step assures that outliers are not affecting them.

Log likelihood

The scaled and robustified deviations $\langle d_i \rangle$ are all treated as standard normal random variables, so the likelihood is:

$$L = \prod_i \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(\frac{-d_i^2}{2\sigma^2}\right) = \prod_i \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-d_i^2}{2}\right) \text{ because } \square = 1$$

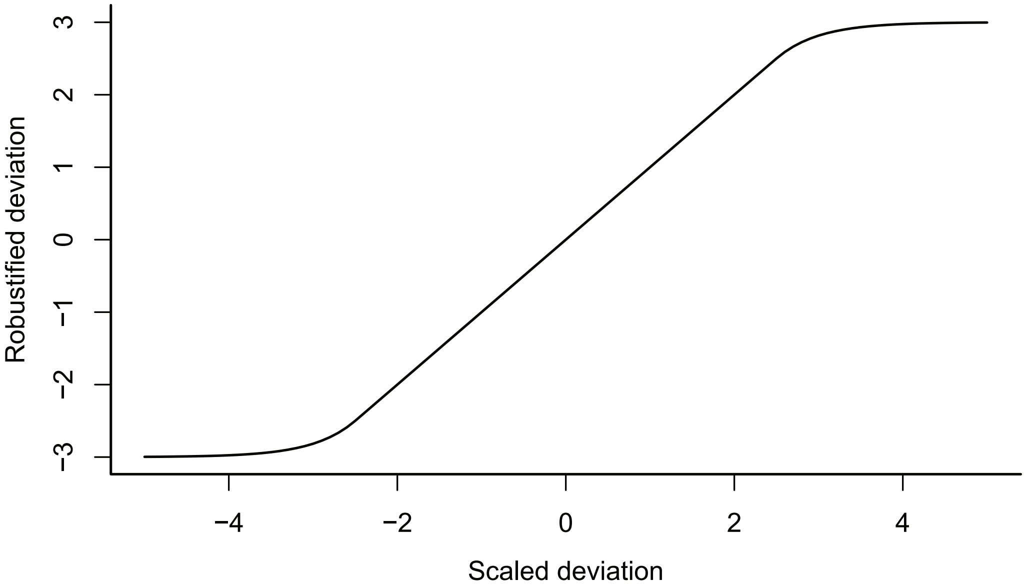


Figure 11. The robustification function. The robust deviation is equal to the scaled deviation up to 2.5 in absolute value. Thereafter the robust deviation increases slowly in absolute value toward an asymptote at 3.

σ_{Sel}

and the log likelihood is $\log L = \text{ constant } + \sum_i -d_i^2/2$. The estimates are located by minimizing $-\log L = \sum_i d_i^2/2$ which is just half the sum of scaled, squared deviations. Taking half the sum rather than just minimizing the sum of squared deviations assures that the log likelihood is scaled properly for estimating the variance-covariance matrix of the parameter estimates from the inverse Hessian matrix.

Penalties

In addition to the sum of squared deviations for the data, the objective function includes some penalty sums of squares for the parameter estimates that serve to control certain features of the estimates. The most important of these is a penalty on the second differences of the selectivity parameters that in effect requires length-specific selectivity to vary smoothly with length or, in other words, prevents the selectivity function from making large changes in direction at any point. Specifically, if $Sel_k, k = 1,...,K$ are the estimated selectivities at the *K* waypoint lengths of a schedule where selectivity is estimated (or possibly fixed), the penalty is calculated as $\sum_{k=3}^{K}\left(Sel_k - 2 \cdot Sel_{k-1} + Sel_{k-2}\right)^2 \big/ \left(2 \cdot \sigma_{Sel}^2\right)$, where σ_{Sel} is an assigned tolerance. In practice a value of 0.025 is sufficient to produce acceptably smooth selectivity schedules. This penalty is typically about 5% of the total sum of squares.

Another penalty is calculated on year-to-year relative changes in commercial catchability. If the estimated catchabilities for the *Y* years of data are $\left\langle CQ_y \right\rangle$, the penalty is calculated as $\sum_{y=2}^{Y}\left(\log CQ_y - \log CQ_{y-1}\right)^2 \big/ \left(2 \cdot \sigma_{CQ}^2\right)$ where σ_{CQ} is an assigned tolerance, presently 0.03, which was found sufficient to prevent wide swings in the estimates. The penalty is not calculated at waypoints where there is a break in commercial catchability, such as at the time of the changeover from J-hooks to C-hooks. This penalty discourages rapid changes in estimated commercial catchability but allows for large cumulative changes over the course of several years, which doubtless occur (Fig. 8). This device makes the commercial CPUE data useful for tracking relative abundance without having to assume that commercial catchability is constant. This penalty is typically about 1% of the total sum of squares.

Two other penalties are calculated to prevent wild estimates of year-class strength or sex ratio among the cohorts present in the stock in the first data year. Wild values of year-class strength can occur because of age smearing, and they are prevented by penalizing large differences in abundance at successive ages in the initial numbers. Wild sex ratios can occur because the initial numbers of females and males at each age in the first year are estimated independently. (Year-classes that recruit in subsequent years are assumed to have an equal sex ratio at age one.) Wild sex ratios are prevented by penalizing deviations from an equal sex ratio through age 8, and large differences in sex ratio at successive ages thereafter. These penalties have no effect on the estimates of present abundance in Areas 2 and 3A where the first data year is 1974, and little effect in Areas 3B and 4. These two penalties together are typically about 5% of the total sum of squares.

Weights

The catch at age data points far outnumber the CPUE data points (survey and commercial) because every year there are 15 or 20 of the former and only one of the latter. As a result the best model fit may be one that does not agree very well with the abundance trends indicated by the CPUE series, especially the survey CPUE series where catchability is held constant. The staff regards the survey CPUE as the most reliable index of long-term changes in relative abundance, and the commercial CPUE as a reliable index of short-term changes in abundance, so we normally weight the CPUE deviations much more heavily than the catch at age deviations in order to obtain fits that agree well enough with the CPUE trends. At present a weight of 10 is used, which gives the catch at age and CPUE deviations roughly equal shares of the total sum of squares. It also provides model predictions that fit both the catch at age data and the CPUE trends reasonably well.

Assessment results

When the annual assessment is done, several candidate models may be considered that differ in how commercial and survey catchability and selectivity are parameterized and how the CPUE data are weighted. Putting in more catchability and selectivity waypoints will generally improve the fit but reduce the precision of the parameter estimates by reducing the number of data points per parameter estimated. The staff catch limit recommendations are normally based on the most parsimonious model that provides acceptable fits in all areas and performs well in retrospective runs. Once such a model has been adopted, the staff attempts to use it in the following year so as to avoid changes in biomass estimates due to changes in model structure or parameterization. But as explained above, there have been frequent changes over the course of the last ten years or so.

Parameterization and weighting of the 2004 assessment model

The 2004 production assessment model had the following features:

(i) The commercial catchability waypoints in Areas 2 and 3A were 1974, 1982, 1984 and every four years thereafter, with interpolation between all waypoints except for 1984 when the C-hook data begin. In Areas 3B and 4 the waypoints were 1996, 2000, and 2004.

(ii) A single length-specific commercial selectivity schedule was estimated for all years in the data series, J-hook as well as C-hook. This was decided after fitting the model with a number of selectivity waypoints and seeing that the commercial selectivity schedules for different periods were all very similar.

(iii) The survey catchability waypoints were only 1974 and 1984, with no interpolation, because we assume that survey catchability has been constant apart from the effect of the change from J-hooks to C-hooks in 1984.

(iv) A single length-specific survey selectivity schedule was estimated for all years, because here too separate schedules estimated for the J-hook and C-hook years were almost the same.

(v) Length-specific bycatch selectivity waypoints in Areas 2 and 3A were 1974 and every five years thereafter; in Areas 3B and 4, 1996 and 2001. Bycatch selectivity was not interpolated between waypoints, so a single schedule applied throughout each five-year period. The five-year periods were chosen because that was the most parsimonious parameterization of bycatch selectivity that still provided satisfactory predictions of bycatch length compositions (Fig. 12). Bycatch selectivities are estimated solely for the purpose of accurately accounting for the removals, so it would be natural to estimate a schedule for each year. The drawback is the large increase in the number of model parameters.

(vi) Commercial and survey CPUE data were given a weight of 10.

Quality of fits

The model predictions of commercial catch at age of females and males are very good (Fig. 13). This is remarkable because all of the predictions are based on a single length-specific commercial selectivity schedule that is used for both sexes in all years, when female length at age was always larger than male length at age and both declined dramatically in the 1990s (Fig. 7). The predictions also navigate the change from surface to break-and-burn age readings in 2002 successfully, and they correctly predict the growing preponderance of females in the catches (Fig. 14). The fits to survey age compositions are not as good but they are quite satisfactory for years since the systematic surveys resumed in 1993 (Fig. 15).

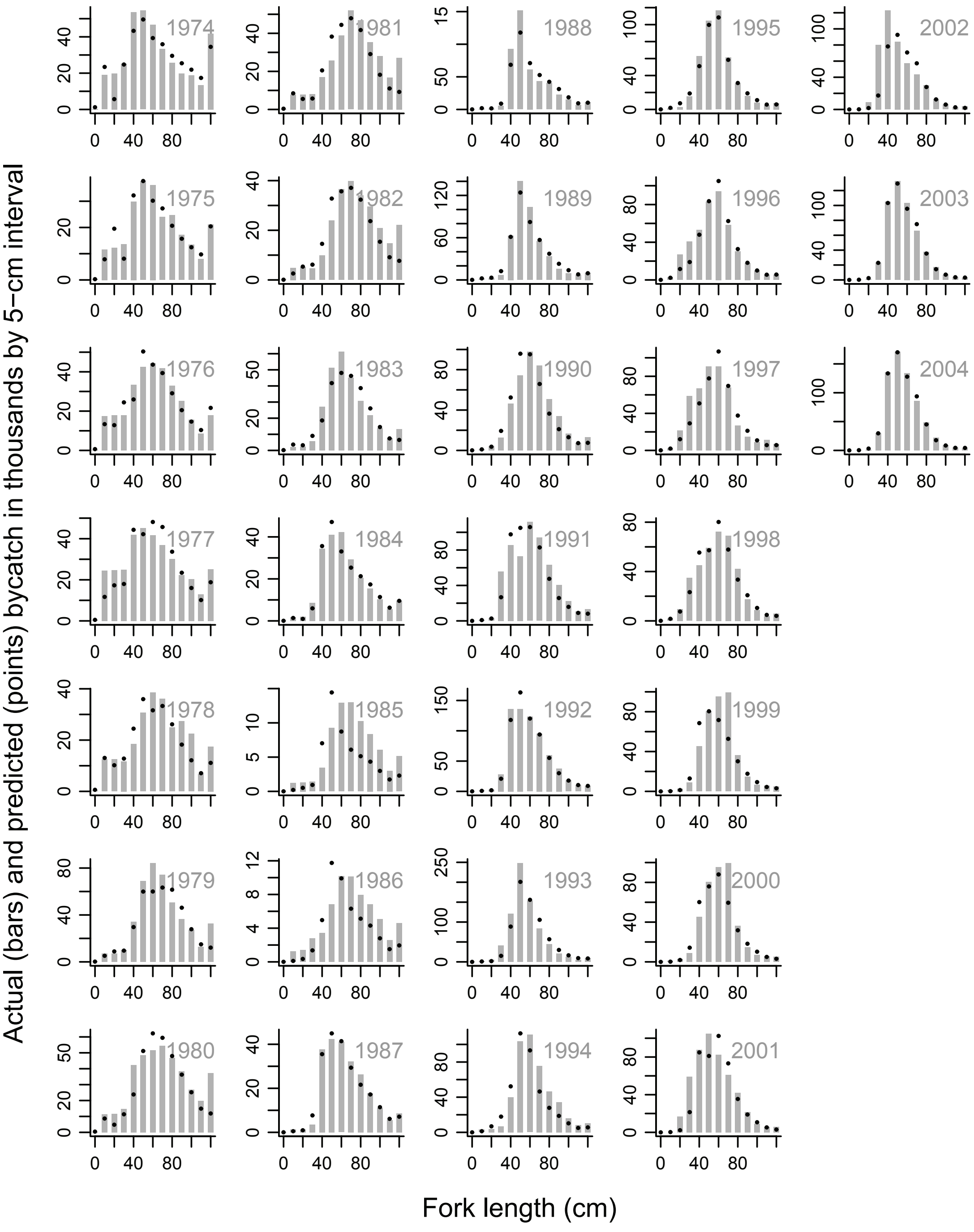


Figure 12. Actual (bars) and predicted (points) bycatch at length in the 2004 Area 3A assessment. An average bycatch selectivity was estimated for each five-year period.

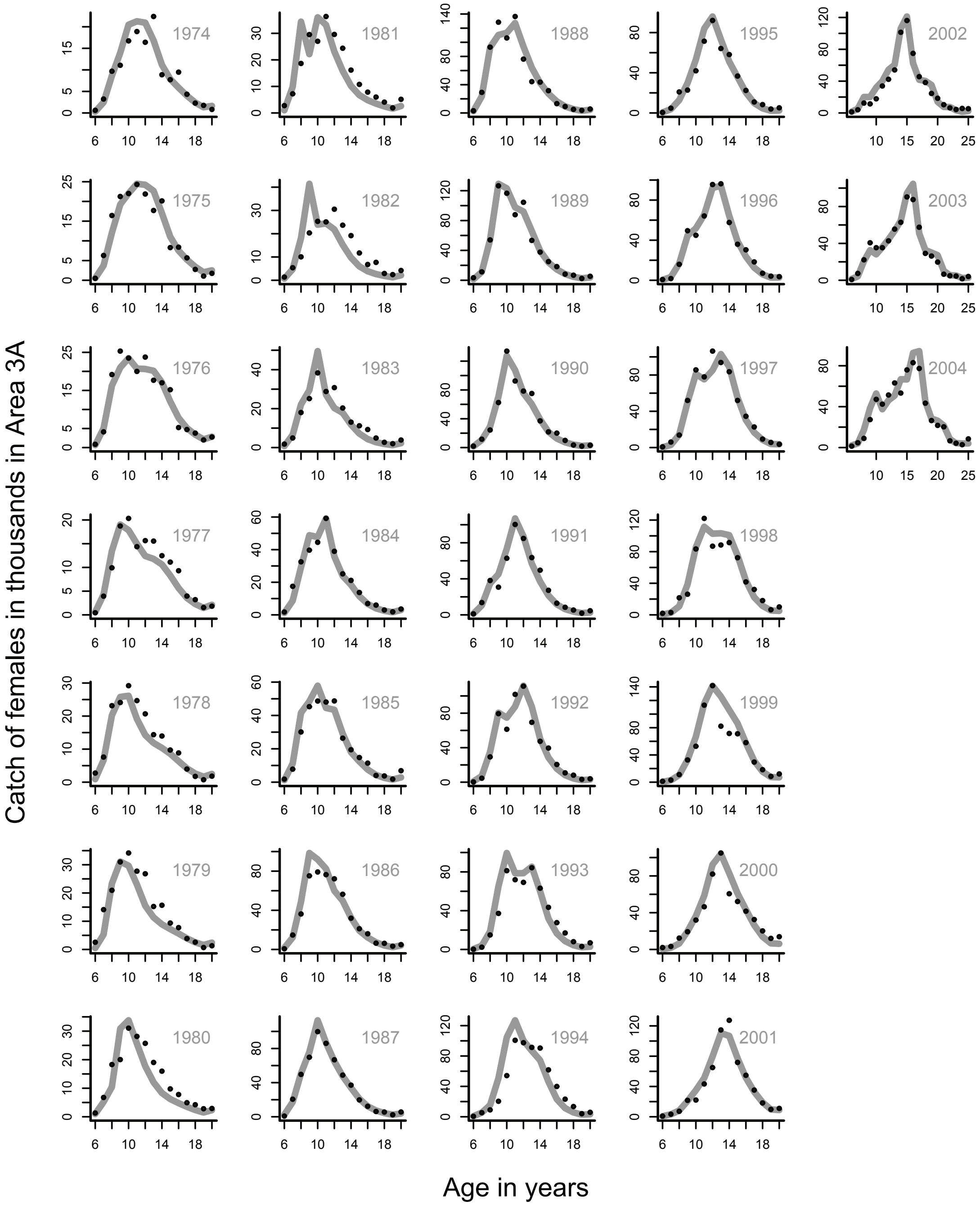


Figure 13a. Observed (points) and predicted (gray lines) commercial catch at age of females in the 2004 assessment in Area 3A.

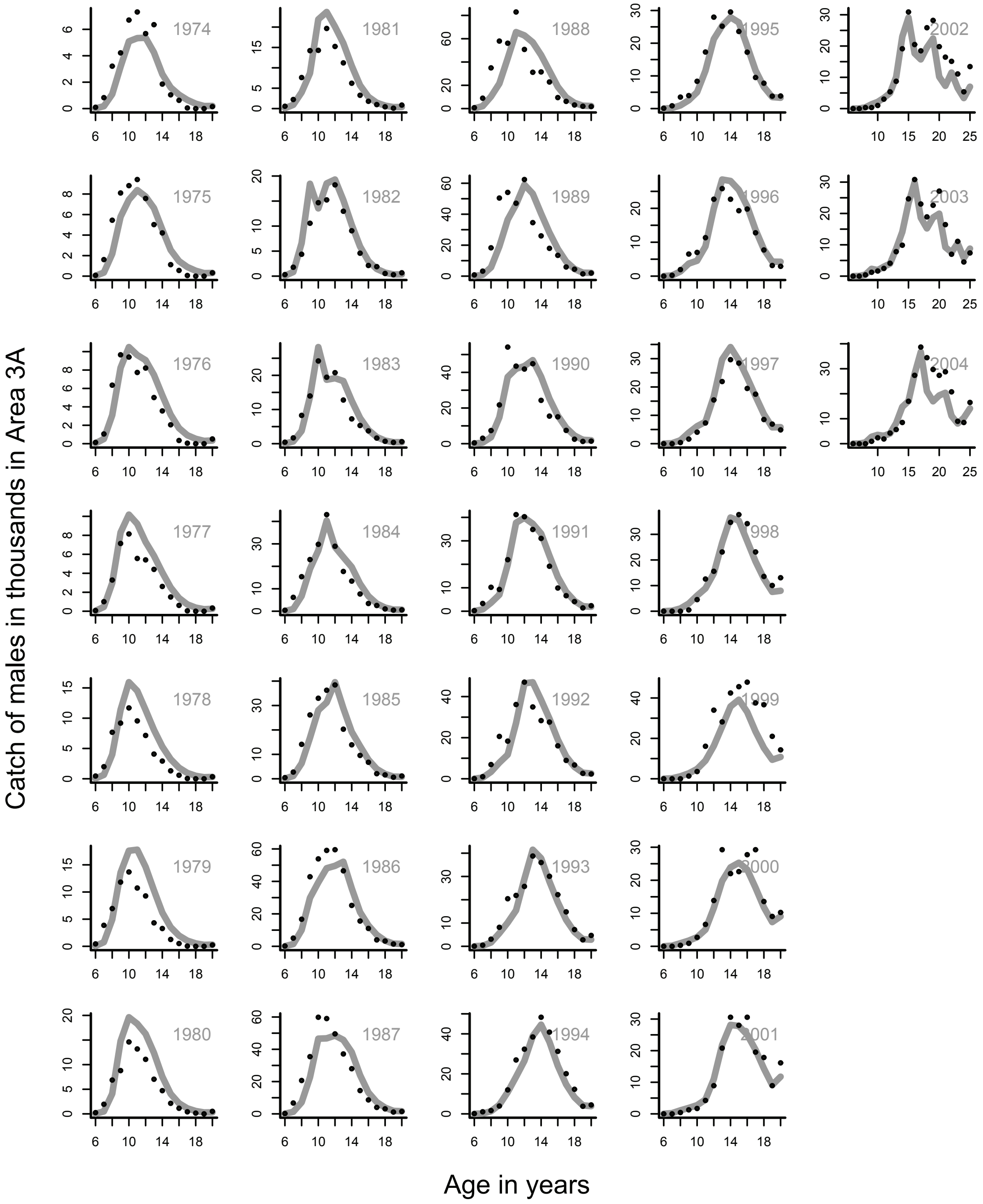


Figure 13b. Observed (points) and predicted (gray lines) commercial catch at age of males in the 2004 assessment in Area 3A.

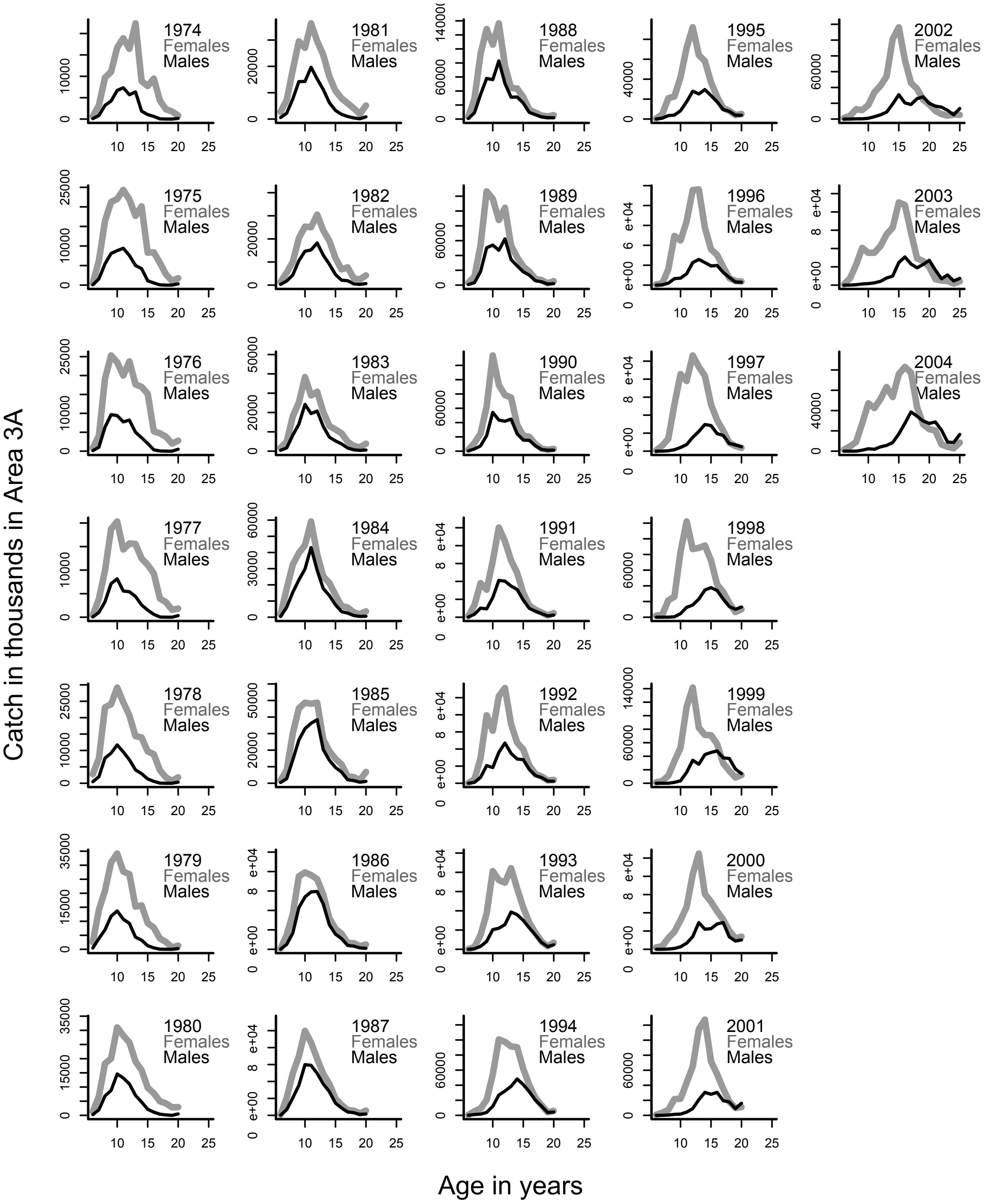


Figure 14. Catch at age of females and males in Area 3A. The values plotted are the data values not the predictions, which track the data very well as shown above.

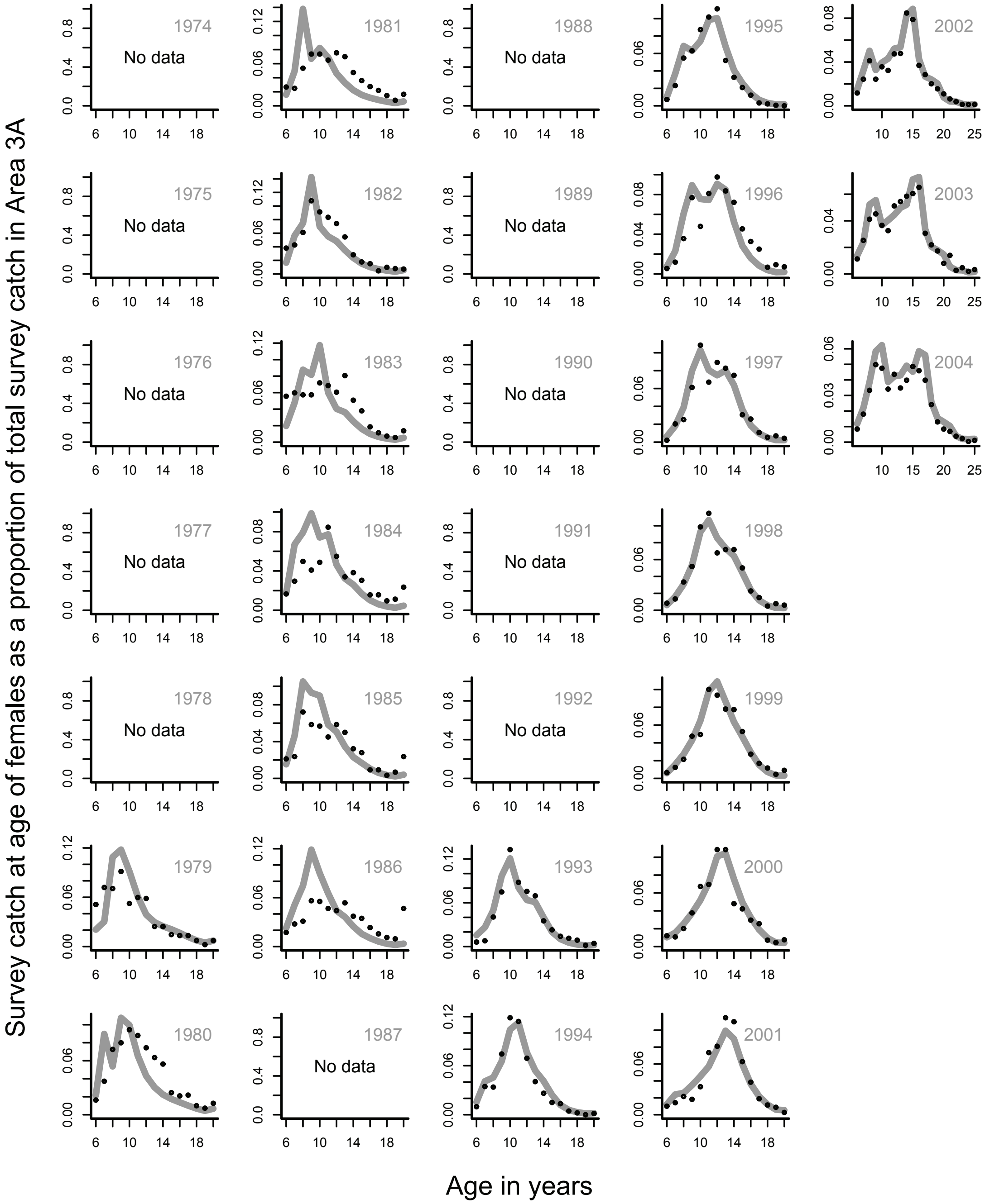


Figure 15a. Observed (points) and predicted (lines) survey catch at age of females as a proportion of total survey catch (including males), in Area 3A. The proportions sum to the proportion female in each year, not to one.

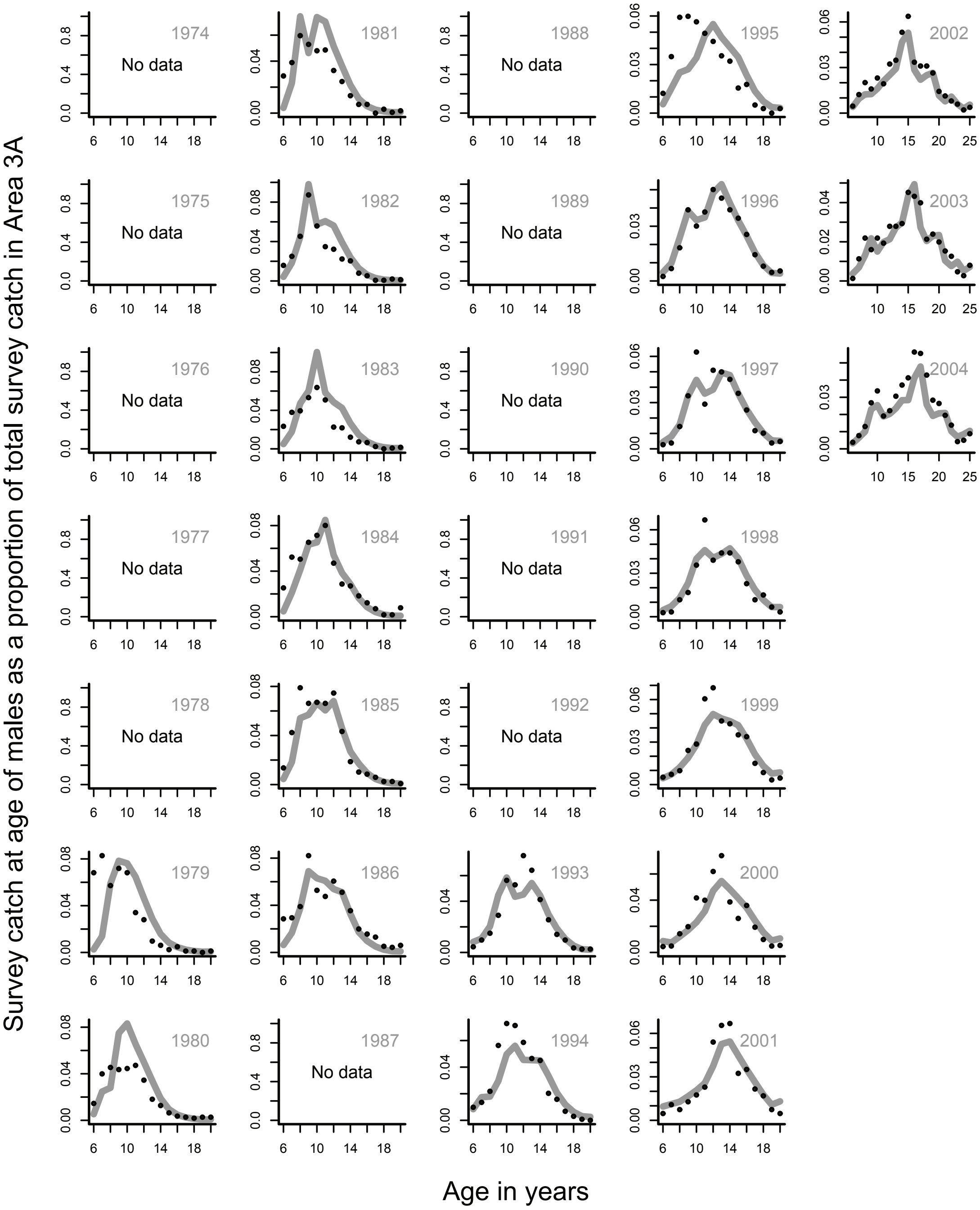


Figure 15b. Observed (points) and predicted (lines) survey catch at age of males as a proportion of total survey catch (including females), in Area 3A. The proportions sum to the proportion male in each year, not to one.

The model CPUE predictions, shown in the upper right panels of Figures 16a-f, also track the observations quite well. (The large jump in all series in 1984 is the result of the change from J-hooks to C-hooks, not a sudden increase in abundance.) That is to be expected in the case of the commercial CPUE because commercial catchability is allowed to vary somewhat over time, but survey catchability is held constant. Because of the heavy weighting of the CPUE series, the model fits are effectively required to match the long-term trend in survey CPUE, and they do, but except for a few years in Areas 2B and 2C the fits also follow the year-to-year trajectory of survey CPUE reasonably well.

Overall the good model fits inspire some confidence that the model is correctly specified and that the procedures used to estimate the commercial catch at age by sex and to allow for the bias and variance of surface age readings, are working properly. Stock trends are therefore probably estimated correctly. The estimates of absolute abundance are conditional on the working value of the natural mortality rate, and are almost certainly high or low by some proportion. (And if the natural mortality has changed over time, the estimated trends are also in error to some extent.)

Retrospective performance

Every year the assessment model is fitted to data going back to 1974 (1996 in Areas 3B, 4A and 4B), so the historical biomass series is re-estimated along with the present biomass. If the model were correctly specified, each year's assessment would reproduce the previous year's estimate of the trajectory of historical biomass with only minor, random differences. The success or failure of a model in this respect is called retrospective performance. Poor retrospective performance is a systematic movement of the estimated biomass trajectory from one year to the next. It can result from trends in model parameters that are not allowed for in the model specification, or from features of the data.

Figure 17a shows the poor retrospective performance of the 2004 model in Area 2B when data are added one year at a time and the same model is fitted to the successive series. An assessment based on data through 1997 shows a large increase in biomass from 1985 to 1997, but subsequent assessments steadily reduce their estimate of the 1997 biomass. According to the 2004 assessment, the 1997 assessment overestimated current biomass by close to 50%. This is poor retrospective performance, but in this case the problem lies in the data. It is now clear that in the mid-1990s there were three anomalously high survey CPUE values in succession in Area 2B (Fig. 16a). At the time, however, those were the current data, and they consistently indicated a greatly increased abundance, which the assessment estimated. As later years' data with lower CPUE values accumulated, the assessment's estimate of abundance at that time was steadily revised downward; hence the retrospective pattern.

Figure 17b shows the retrospective performance of the assessment in Area 3A, where there were no rogue survey CPUE values, and it is acceptable. The estimated biomass trajectory does drop a ways from the 1997 to 2002 assessments but then come back up. There is not a sustained movement either way, and the estimate of biomass in 1997 is quite close in the 1997 and 2004 assessments.

Variance estimates

If the model and the error structure are correctly specified, the inverse Hessian matrix evaluated at the maximum likelihood estimate is a good estimate of the variance-covariance matrix of the parameter estimates. In fisheries assessments the model is always a gross simplification of the stock and fishery, and thus misspecified to some extent, and the data are always over-dispersed (meaning that there are more extreme values in the data than would be expected from the assumed sampling distribution). For both reasons the usual Hessian-based estimates of variance are always too low, as are alternative methods of variance estimation like bootstrapping and Markov chain Monte Carlo sampling (Punt and Butterworth 1993). For lack of anything better, such estimates are often reported anyway.

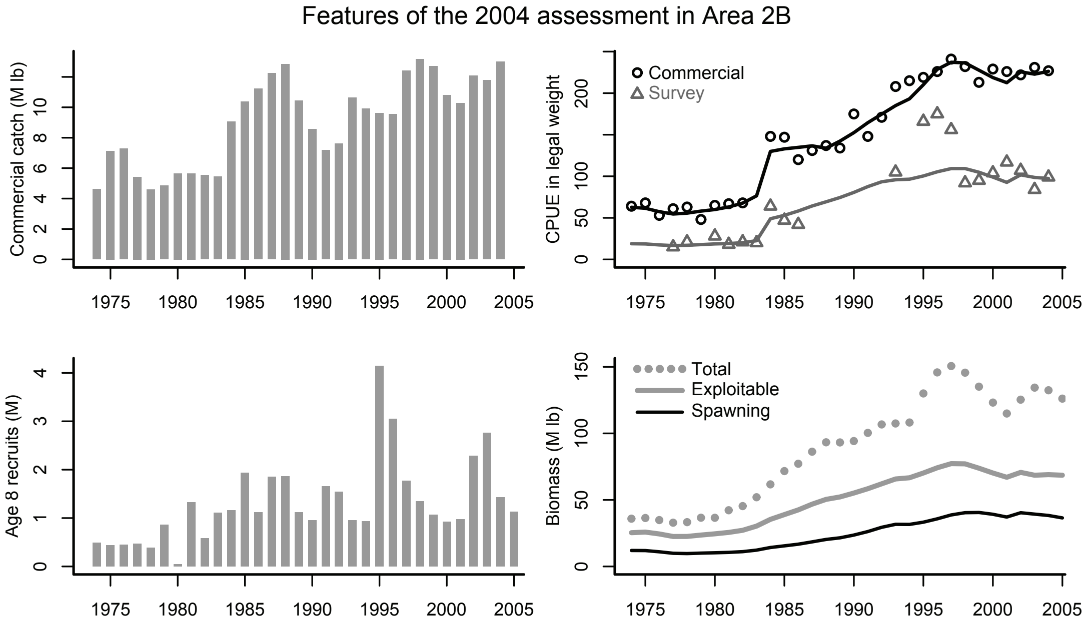


Figure 16a. Area 2B stock trends as shown by the 2004 assessment. In the figure at upper right, the points are observed CPUE values and the lines are model predictions. The recruitment and biomass series are model estimates.

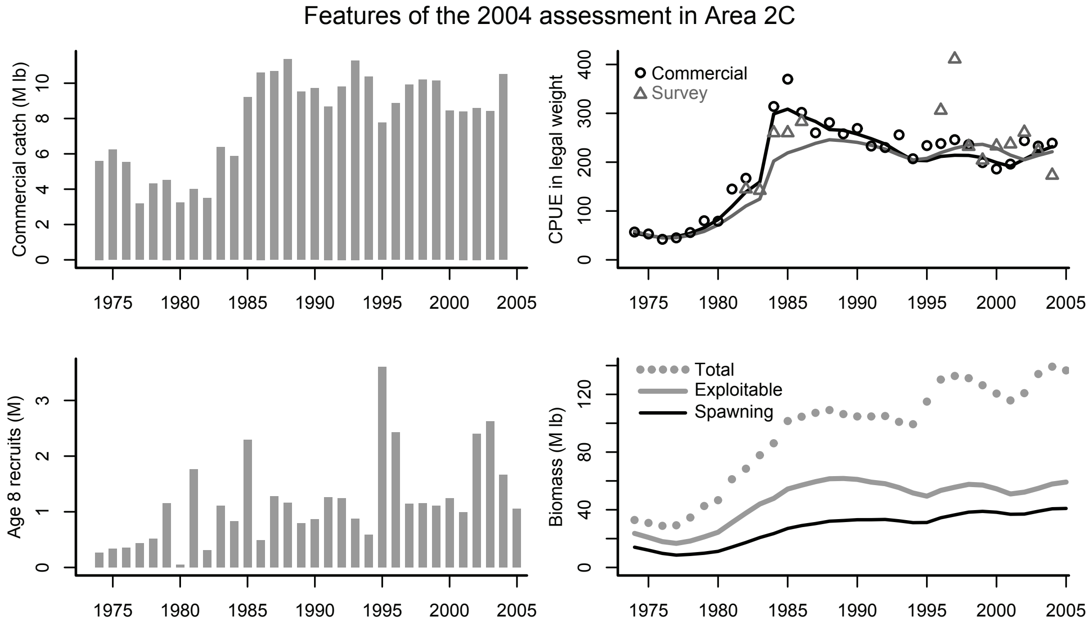


Figure 16b. Area 2C stock trends.

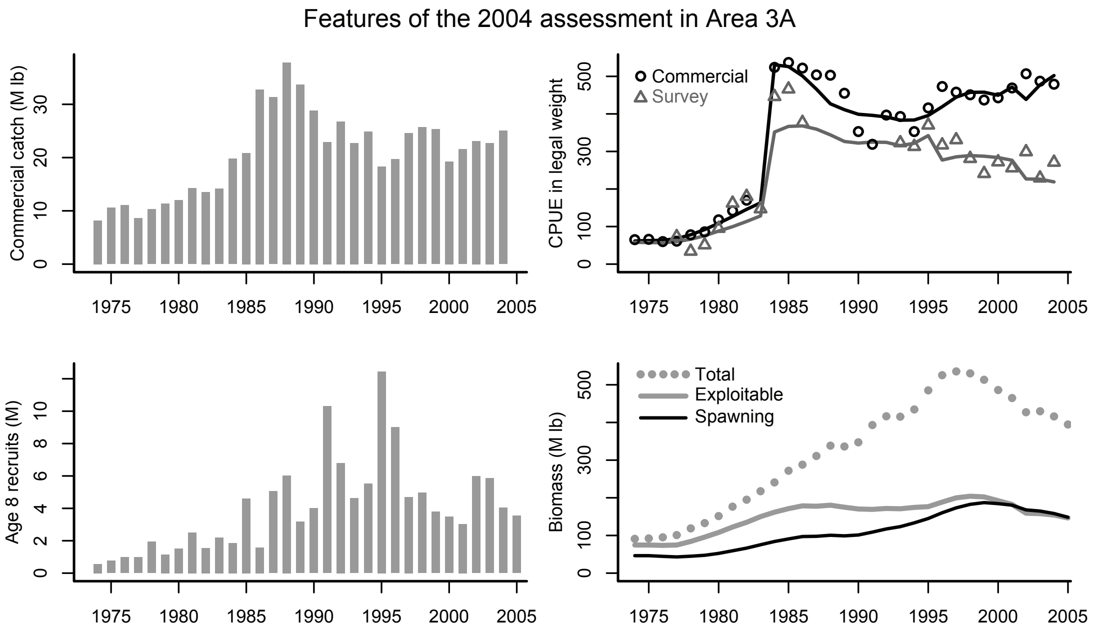


Figure 16c. Area 3A stock trends.

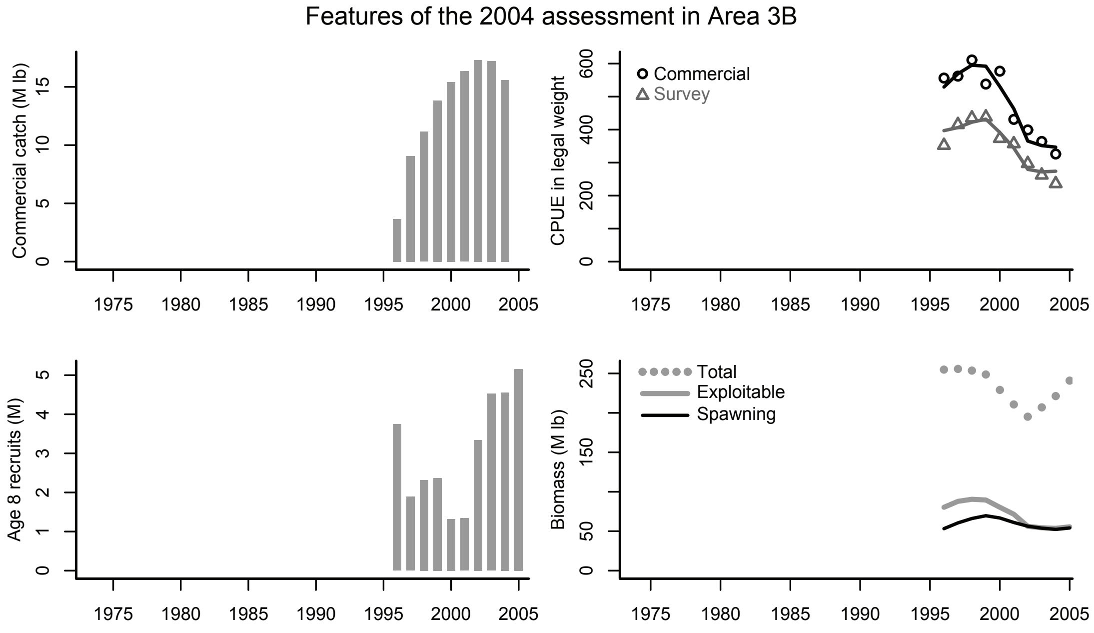


Figure 16d. Area 3B stock trends.

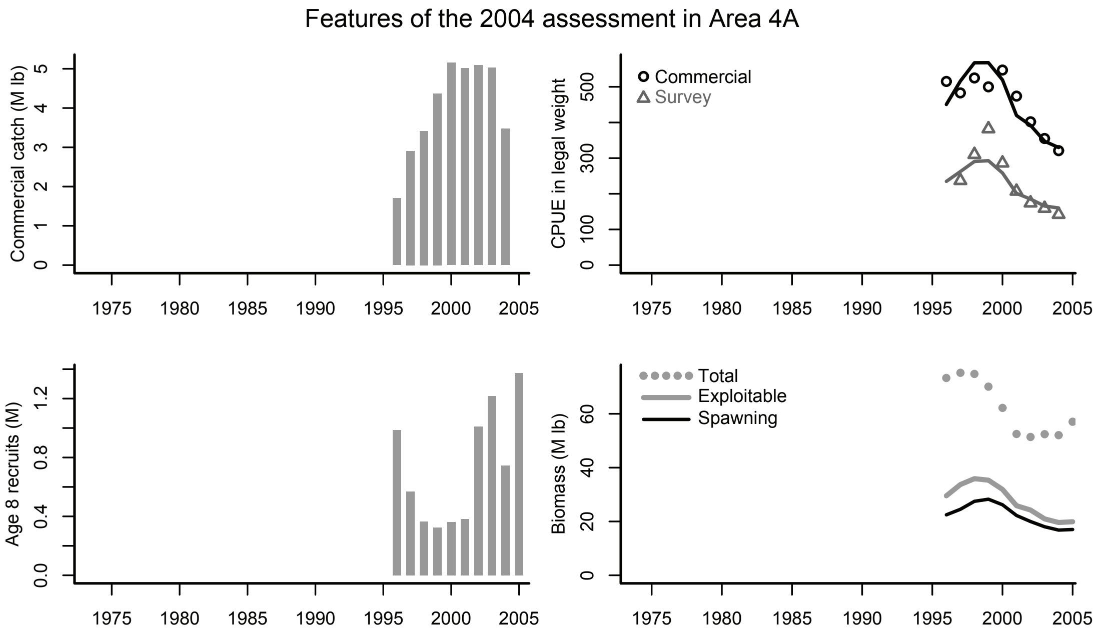

Features of the 2004 assessment in Area 4A

Figure 16e. Area 4A stock trends.

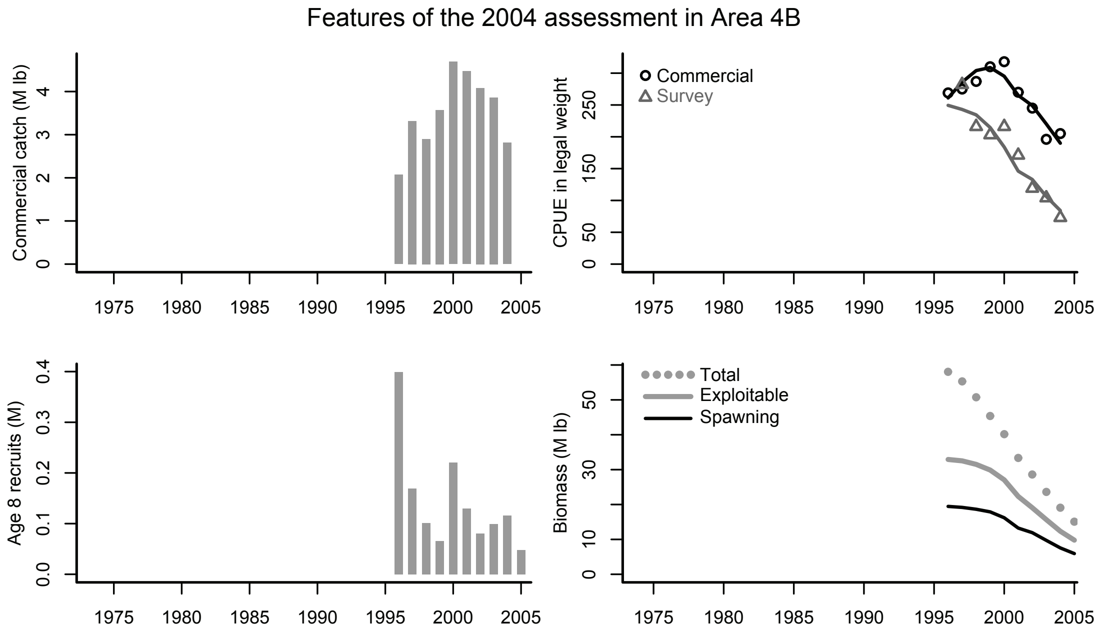

Features of the 2004 assessment in Area 4B

Figure 16f. Area 4B stock trends.

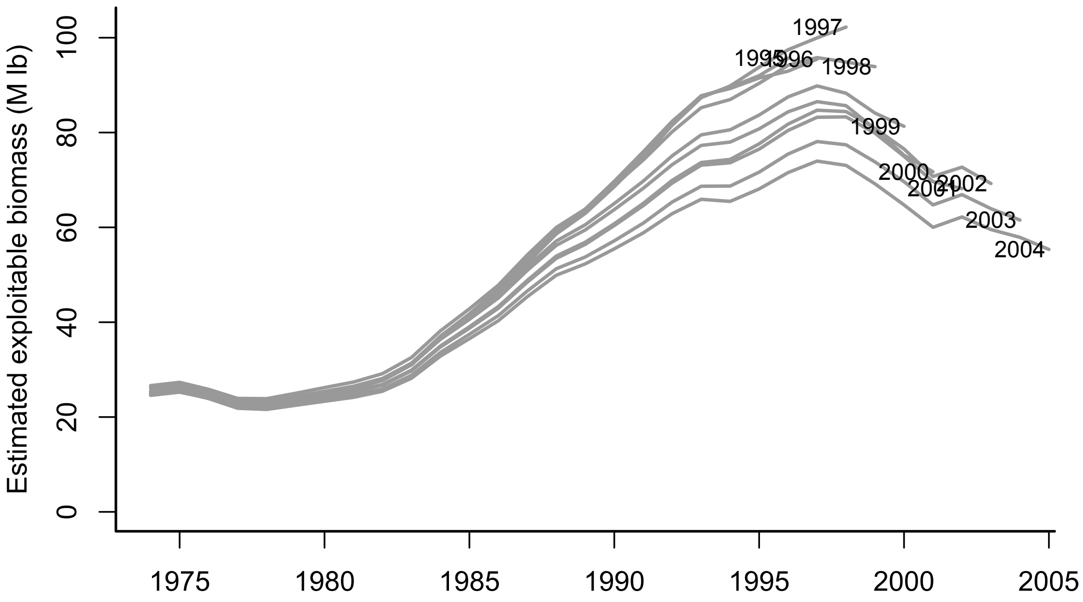


Figure 17a. Retrospective performance of the 2004 assessment in Area 2B. Each plotted line shows the biomass trajectory

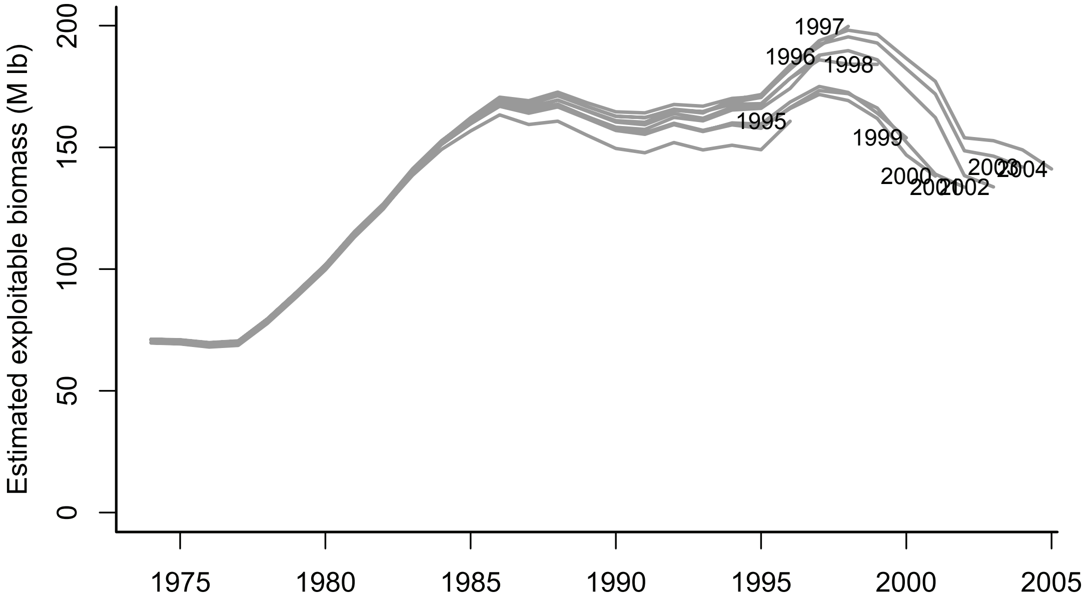


Figure 17b. Retrospective performance of the 2004 assessment in Area 3A.

In the case of the halibut assessment, the standard Hessian-based estimate of the coefficient of variation (CV) of the present biomass is less than 5%, because the heavy weighting of the CPUE series inflates the log likelihood. From a statistical viewpoint, the weighting amounts to saying that we have a lot of CPUE data, whereas in fact we have a small amount that we choose to rely on heavily. In unweighted fits of the model, the CV of the present biomass estimate is about 10%, and even that is low because the log likelihood includes some double counting. For example, the sum of squares includes the total catch at age as well as the female catch at age and the male catch at age; likewise the CPUE values in both number and weight. If we allow for that by dividing the sum of squares by 4 instead 2 when computing the log likelihood, the CV of the present biomass estimate in unweighted fits increases to 15%.

An alternative estimate can be made from the retrospective performance of the fit. In particular, the observed variance of the estimate of present biomass must overstate the variance of the estimate itself because it also includes any variance in the true biomass. Specifically, if B_y is the true biomass in year *y* and $\hat{B}_y$ is the estimate of it in that year's assessment, then the variance of $\hat{B}_y$ among years is $V\left(\hat{B}_y\right) = V\left(B_y\right) + V\left(\hat{B}_y \mid B_y\right)$. Over the last ten years, the CV of the present biomass estimate has been about 15% in Areas 2C and 3A, and about 20% in Area 2B (owing to the rogue survey CPUE values).

Considering both ways of estimating variance, we believe that a CV of 10-15% is a reasonable value for the present biomass estimate, but as the Area 2B example shows the fits are capable of much larger excursions which can persist for years.

Estimates of present female spawning biomass in Areas 2B, 2C, and 3A relative to 1974

The Commission's paramount management objective is to maintain a healthy level of spawning biomass, meaning a level above the historical minimum that last occurred in the mid-1970s. Although low, this spawning stock nevertheless produced average or better year-classes. One of the main reasons for implementing a sex-specific assessment was to obtain direct estimates of female mortality rates and female spawning biomass. We now have those estimates, and fortunately they show that female spawning biomass is 3-4 times what it was in the mid-1970s (Figs. 16a-c). So on that score the stock is in good shape.

The numbers of fish aged 8 and older are now 5-10 times what they were in 1974, but their total biomass is only 3-5 times the 1974 level, and exploitable biomass (computed with length-specific commercial selectivities as explained below) only 2-3 times. The difference between the large increase in numbers and the more modest increase in biomass results from the dramatic decline in size at age and therefore commercial selectivity that has occurred over the last fifteen years. A significant part of the age 8+ biomass now consists of males that never get large enough to be caught in any numbers, as shown by their near disappearance from commercial catches in Area 3A (Fig. 6b). Looked at another way, in 1974 a large fraction of the total age 8+ biomass was exploitable; now that fraction is much smaller.

Length-specific and age-specific commercial selectivities

As in previous length-specific model fits (in the 1990s), commercial selectivity is estimated to be higher in Area 2B than in Area 3A, with Area 2C intermediate (Fig. 18). The estimates for Areas 3B, 4A, and 4B are similar to the Area 2C estimates.

Because length-specific commercial selectivity appears to have been the same for the last thirty years while mean length at age has declined greatly over the last fifteen years, age-specific commercial selectivity has also declined greatly over the last fifteen years (Fig. 19). And because males in the modal age range (10-15) were less vulnerable to begin with, the relative decline in age-specific selectivity of males has been greater than that of females. In Area 3A, males reached full vulnerability by age 15 in the 1970s and 1980s; now even the oldest males are only about 50% vulnerable, while the oldest females are still fully vulnerable. The same sort of change has occurred elsewhere. Females always sustained higher fishing mortality rates than males because they were larger, but twenty years ago females and males both reached the size of full vulnerability at some point. Males no longer reach that point, so an even larger share of fishing mortality is falling on the females.

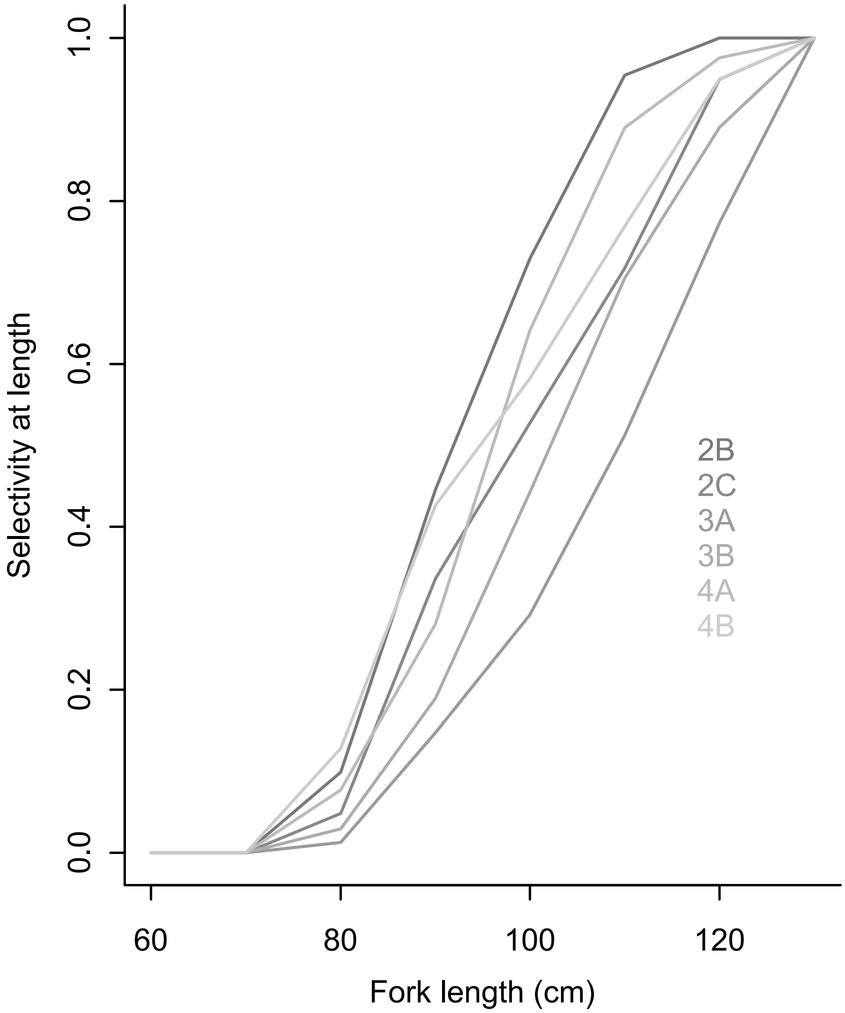


Figure 18. Estimated length-specific commercial selectivity. The topmost line is Area 2B. The bottom line is Area 3A, and the other Alaska areas are clustered in the middle.

Calculation of exploitable biomass

The exploitable biomass of an age/sex group is the product of surviving number, commercial selectivity of the group, and mean weight of the group in commercial landings. The staff has always tried to use the same commercial selectivity schedule in all areas so as to report a single measure of biomass. A common schedule is used for all Alaska areas (approximately the Area 2C estimate), but in Area 2B the local estimate is used because it is substantially higher than all of the Alaska schedules.

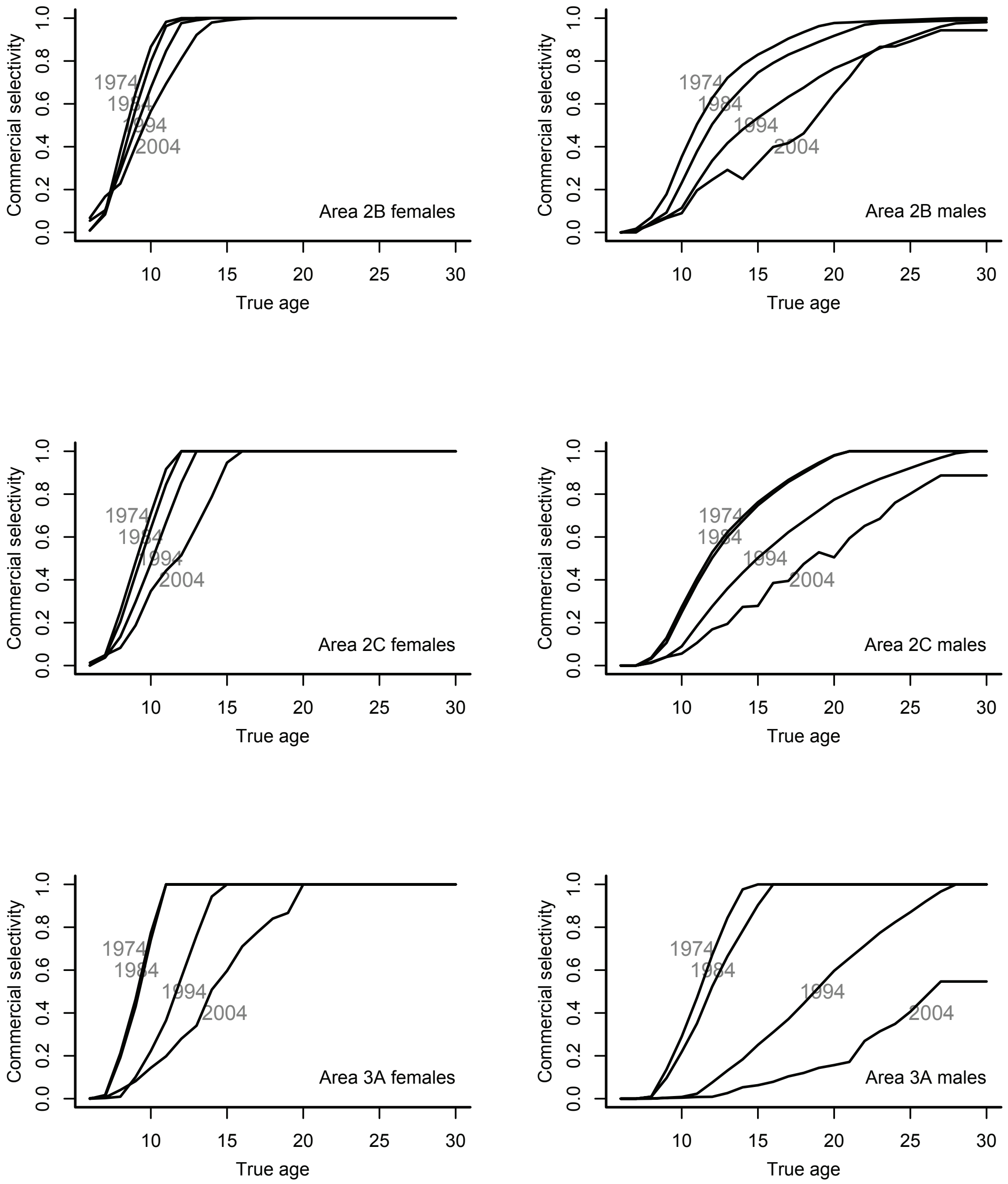


Figure 19. The downward drift of age-specific commercial selectivities over time due to constant length-specific selectivity and declining size at age, plotted by area and sex.

Estimation of exploitable biomass in Areas 2A and 4CDE

Areas 2A and 4CDE present special problems. In Area 2A there are gaps in the commercial data, a history of changes in the commercial and treaty fisheries, sparse survey data, and a large bycatch removal for which we have no length data. Also in Area 4CDE setline survey data are scanty and, like the fishery, limited to the shelf edge. In both areas we lack the data series used in the standard assessment in other areas and so have resorted to an ad hoc procedure to estimate present biomass.

Exploitable biomass in Area 2A is calculated as a proportion of the Area 2B analytical estimate. The proportion used is the ratio of setline survey CPUE's (three-year running mean) weighted by bottom areas:

$$\frac{(\text{2A biomass})}{(\text{2B biomass})} = \frac{(\text{2A CPUE}) \times (\text{2A bottom area})}{(\text{2B CPUE}) \times (\text{2B bottom area})}$$

The idea here is that survey CPUE is an index of density and multiplying it by the total bottom area gives an index of total biomass. The value of the scaling proportion in 2004 was 12%. In the same way, exploitable biomass in Area 4CDE is calculated as a proportion (160% in 2004) of the Area 4A analytical biomass estimate.

The bottom areas used for these calculations (Table 2) are 0-300 fathoms, the upper limit chosen because very few halibut are caught below 300 fathoms in summer.

Estimated abundance 1935-1973

The modern Pacific halibut stock assessment model produces abundance estimates for the years 1974-present (1996-present in Areas 3B, 4A, and 4B). The modern assessment begins in 1974 because stock biomass reached the historical minimum at that time, and it is of interest to see the comparison between the estimates of present biomass and that reference point. It would not be possible to start the modern assessment any earlier because the survey data only go back to the late 1970s.

The Pacific halibut fishery dates to the late 1890s, and estimates of abundance before 1974 are of interest for a variety of reasons, including evaluating stock-recruitment and environment-recruitment relationships, determining productivity of the stock, detecting evidence of density-dependence and establishing minimum and maximum biomass levels. Catch and effort data from the halibut fishery have been collected since at least the mid-1920s, but the size and age data required for catch at age modeling were first collected in 1935. Data adequate for fitting an age-structured assessment model are available for IPHC regulatory areas 2B, 2C, and 3A. In the western part of the Commission area fishing was light and spotty before the mid-1990s, and there are some large gaps in the data series.

Table 2. Bottom areas of IPHC regulatory areas (0-300 fm), in square nautical miles.

Area	Bottom area
Area 2A in total	12100
North of Pt. Chehalis at 46º53′ N (“2A-1”)	4077
Remainder (“2A-2”)	8023
Area 2B in total	28100
North of Vacouver Island (50º45′N)	22074
Remainder (W. coast Vancouver Island)	6026
Area 2C	15000
Area 3A	49500
Area 3B	30200
Area 4A	18500
Area 4B	16200
Area 4CDE (to about 60º N)	120000
Area 4C	9600
Area 4D edge (75-300 fm)	5000

The model that is fitted to estimate historical abundance for the period 1935-1973 is simpler than that used for the modern assessment. Absent survey data, the sexes are not distinguished. Only commercial CPUE is available as an abundance index. The historical assessment is joined to the modern assessment by forcing the numbers at age (for sexes combined) to match the "smeared" numbers at age in 1974 in the modern assessment. Because commercial catchability is allowed to drift, the model fit to the early years' data is essentially a VPA (Virtual Population Analysis) with the terminal values fixed, so the historical estimates are entirely determined by the catch at age data; they have no variance to speak of. The historical and modern assessments together provide a seamless time series of recruitment and biomass estimates over the entire time period 1935-2005 for Areas 2B, 2C, and 3A.

Input data

The basic input data for the historical assessment model are removals at age, weight at age, and effort data. Removal data consist of commercial catch, bycatch, sport catch, personal use, and wastage. The following input data can be found in tables in Hare (2001). A brief description of the data sources follows.

Catch at age

The commercial catch at age data were taken from a recent documentation of IPHC sampling protocols (Clark et al. 2000). These data are for ages 6 to 19 with a plus group for ages 20 and older. (Fish younger than age six are rare in the catch except for Area 2B. The number of such fish that would have survived to age 6 if not caught is added to the assessment estimate of abundance at age 6 when calculating estimates of recruitment.) Catch numbers at age are decremented using an annual natural mortality rate of 0.15 to estimate number of absent six year olds. Strictly speaking, recruitment losses should be distributed among downstream areas using a migration schedule as is done for juvenile bycatch losses (Clark and Hare 1998). That is not done here because only in Area 2B is juvenile catch of any significance and those fish would almost all recruit to Area 2B. Lost recruitment is simply added to estimated numbers at age in the area where the juveniles were captured.

The effects of bycatch are dealt with in a direct and comprehensive manner. Estimates of halibut bycatch mortality (i.e., mortality of halibut in fisheries other than the directed setline fishery) begin in 1962 (Williams et al. 1989); historical length-frequency distributions of bycatch have been assembled recently (Hare et al. 2004). Using a length-age key, the bycatch mortality can be divided into adult (age 8+) and sublegal mortality (see Clark 2000 for details).

Adult and juvenile bycatch mortality are incorporated differently. Adult bycatch mortality is added to the commercial removals and effort data are expanded for the increased removals (see below for methodology). Juvenile bycatch is added to the estimate of age-six recruits in the manner described for commercial catch of juvenile fish. Because the sublegal halibut are in the process of migrating, they are often captured in areas "upstream" from where they would have recruited. The migration model of Clark and Hare (1998) is used to assign recruitment loss by area (using the "Intermediate" model); i.e., lost recruitment is added to the number of six-year- old fish in the area they would have recruited to rather than the area where they were captured (though they are to a large measure the same area).

Three other forms of removal from the halibut population—sport catch, personal use and wastage—are assumed to be negligible prior to 1974.

Weight at age

Weight at age has been estimated in a variety of ways over time. The preferred method is to collect fork length samples and estimate weight using a length-weight relationship that has held up well over time (Clark 1992a). However, between 1963 and 1990, fork length was not measured in the field but estimated from otolith measurements (radius from 1963-1967, length from 1968-1977, and weight from 1978-1990). The weight at age data used here for this period were taken from Clark et al. (2000).

Effort

The effort data used here were taken from IPHC Technical Report No. 14 (Myhre et al. 1977) which was re-keyed specifically for this purpose. To be consistent with the effort data used in the modern stock assessment, the values from that report have all been divided by a factor of 2.2, reflecting the difference in fishing power between "J" hooks used before 1983 and "C" hooks used since.

Assessment model

A modified version of the CAGEAN model (Deriso et al. 1985) was used to produce historical estimates of abundance. Prior to fitting, commercial catch at age in number and effort data for each year were expanded to account for bycatch as follows (the temporal subscript is eliminated here for clarity purposes only):

$$C_1 = \sum_{6}^{20+} C_a$$

$$\bar{w} = \frac{\sum_{6}^{20+} \left(C_a \cdot w_a\right)}{C_1}$$

$$C_2 = C_1 + \frac{B}{\bar{w}}$$

$$C'_a = C_a \cdot \frac{C_2}{C_1}$$

$$E' = E \cdot \frac{C_2}{C_1}$$

where C_a is commercial catch at age in numbers, C'_a is expanded catch numbers, B is weight of adult bycatch mortality and E and E' and original and expanded effort numbers, respectively. The operational equations are:

$$C'_{t,a} = N_{t,a} \frac{F_t \cdot Sel_{t,a}}{F_t \cdot Sel_{t,a} + M}\left(1 - \exp\left(-\left(F_t \cdot Sel_{t,a} + M\right)\right)\right)$$

$$N_{t+1,a+1} = N_{t,a} \exp\left(-\left(F_t \cdot Sel_{a,t} + M\right)\right) \qquad a=7, ..., 19$$

$$N_{\mathrm{t}+1,20} = N_{t,19} \exp\left(-\left(F_t \cdot Sel_{19,t} + M\right)\right) + N_{t,20} \exp\left(-\left(F_t \cdot Sel_{20,t} + M\right)\right) \qquad a=20+$$

$$SpBio_{\mathrm{t}} = \sum_{\mathrm{a}=1}^{20+} N_{t,a} \cdot w_{t,a} \cdot Mat_a$$

$$R_t = N_{t,6} + BycatchLosses$$

where C is catch, N is numbers, F is full recruitment fishing mortality, Sel is selectivity, $SpBio$ is spawning biomass, w is weight, Mat is maturity, M is natural mortality and R is age-six recruits. The subscript t indexes time and a indexes age. Recruitments are freely estimated within the model.

Catchability and selectivity

Selectivity is assumed to be a length-specific process. Selectivity was assumed to be 0.0 for lengths less than 60 cm and 1.0 at lengths greater than or equal to 120. Intermediate (monotonically increasing) values Sel_l were estimated at 10 cm intervals with interpolation used to complete the selectivity schedule. Catchability was allowed to drift over time and was implemented as a constrained random walk:

$$\ln(q_{t+1}) = \ln(q_t) + {}_q\varepsilon_t$$

The constraint was implemented in the form of a lambda (inverse variance) in the minimization function.

Parameters

For each area, a total of 137 parameters was estimated:

Parameter		Years	Ages	Number
Initial abundance	$N_{1935,a}$	1935	7-20+	14
Recruitment	$N_{6,t}$	1935-1973	6	39
Full recruitment fishing mortality	F_t	1935-1973	16+	39
Catchability	q_{1935} and ${}_q\varepsilon_t$	1935-1973		39
Selectivity	Sel_l	1935-1973		6

Objective function

Parameter estimates were obtained by fitting the model to observations of catch and effort. The variance associated with effort was taken to be half that associated with catch at age (Deriso and Quinn 1985). The objective function is the sum of the weighted residual sum of square (RSS) terms:

$$RSS(C) = \lambda_C \sum_a \sum_t \left(\ln C_{t,a}^{obs} - \ln C_{t,a}\right)^2 \qquad \lambda_C = 1.0$$

$$RSS(E) = \lambda_E \sum_t \left[\ln F_t - \left(\ln q_t + \ln E_t^{obs}\right)\right]^2 \qquad \lambda_E = 0.5$$

The negative log likelihood –ln L, ignoring constant terms, is

$$-\ln L = 0.5 n_{obs} \ln\left(RSS(C) + RSS(E)\right)$$

where n_{obs} is the total number of observations. For each regulatory area there are 15 age groups and 39 years of catch at age data and 39 observations of fishing effort, for a total of 624 observations.

The objective function also includes three penalty terms: one to constrain the amount of drift allowed in catchability (q_t), another to penalize abrupt changes in selectivity (Sel) and a third on differences between the projected numbers at age (N_a) in 1974 and the corresponding stock assessment estimates (N'_a). The catchability and selectivity standard deviations were set at 0.05 (similar to values used in the modern assessment) while the numbers at age standard deviation was set much lower at a value of 0.01 to insure the final numbers at age match those generated by the modern assessment for the 1974-present period

$$
\begin{aligned}
PSS(Q) &= \sum_{t} {}_{q}\varepsilon_t^{\,2} \Big/ \left(2 \cdot \sigma_q^2\right) \\
PSS(S) &= \sum_{k=3}^{K} \left(Sel_k - 2 \cdot Sel_{k-1} + Sel_{k-2}\right)^2 \Big/ \left(2 \cdot \sigma_{Sel}^2\right) \\
PSS(N_a) &= \sum_{a=7}^{20+} \left(N_a - N'_a\right)^2 \Big/ \left(2 \cdot \sigma_{N_a}^2\right) \\
penalties &= PSS(Q) + PSS(S) + PSS(N_a)
\end{aligned}
$$

The objective function that was minimized is:

$$OF = -\ln L + penalties = 0.5 n_{obs} \ln\big(RSS(C) + RSS(E)\big) + PSS(Q) + PSS(S) + PSS(N_a)$$

Model output

For the purposes of the harvest rate analysis, the assessment provides two outputs of interest: long term estimates of recruitment and biomass. The recruitment estimates are straightforward and are illustrated in Figure 20. (All of these estimates of recruitment at age 6 are adjusted for bycatch mortality before age 6.) As detailed in Clark et al. (1999) and Clark and Hare (2002), the most notable aspect of the recruitment time series is the appearance of alternating recruitment "regimes" of 15-30 years duration. A productive regime occurred from at least 1935 until around the mid 1940s, followed by a relatively unproductive regime that lasted until the mid 1970s. Recruitment from spawning in the late 1970s to the present has been at a very high level. The dynamics of recruitment are discussed more fully below in the "Population dynamics" section.

Establishing a long term, consistent estimate of biomass is more problematic. As noted, the early period assessment is not sex specific and there are no fishery-independent estimates of weight at age. Without the sex data, no long-term estimate of female spawning biomass can be made. The change in size limit that occurred in 1974 also precludes estimation of a consistent exploitable biomass time series. The most consistent time series is a measure of total biomass for halibut age 10 and over. Total biomass is the product of numbers at age and weight at age. For the period prior to 1974, there is only the commercial weight at age data. For 1974 and afterwards, estimated survey weight at age is used to compute total biomass. However, the estimated weight at age for ages 6-9 are very different between the two time series, likely due to differences in selectivity. By limiting the total biomass summation to ages 10 and older the time series show a very smooth transition between the two time periods (Fig. 21).

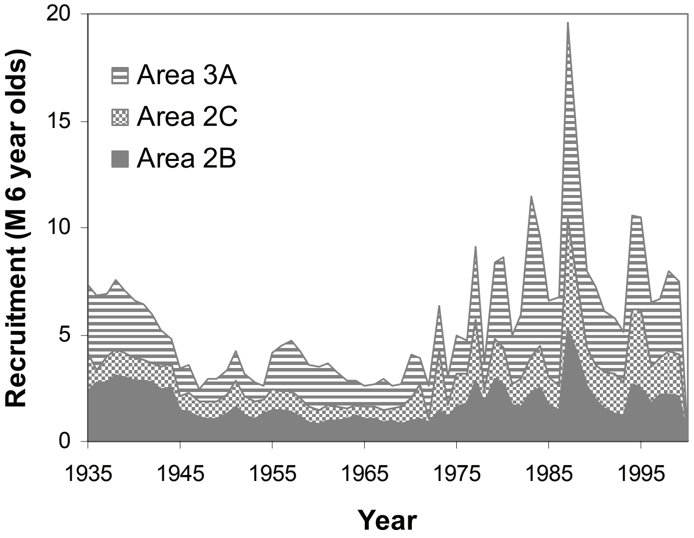


Figure 20. Trend in recruitment of age-six halibut for IPHC Regulatory Areas 2B, 2C, and 3A. The year plotted is the year of spawning.

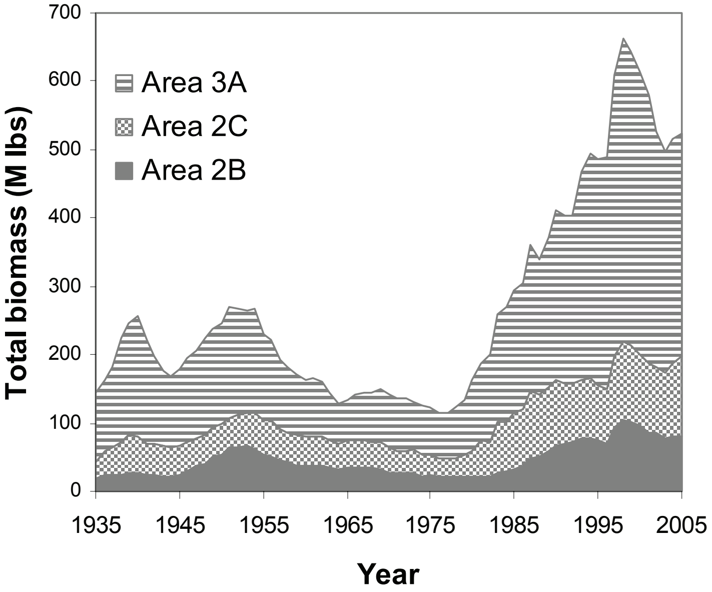


Figure 21. Trend in total biomass (age 10+) of halibut for IPHC Regulatory areas 2B, 2C and 3A.

Population dynamics

The most important aspects of halibut life history as they apply to determining an appropriate harvest strategy are the dynamics of recruitment, growth and maturity. The productivity of halibut depends on its rates of recruitment and growth, both of which have varied greatly over the last 70 years. The pronounced change in size at age also has the potential to affect the maturity and egg production of female halibut. In their analysis of recruitment and growth dynamics, conducted external to the assessment, Clark and Hare (2002) concluded that recruitment was likely environmentally driven (at least within the range of observed spawning stock sizes) while growth was a density dependent process, most closely linked to the number of adult (age 10+) halibut in the population. At the time of their analysis, the halibut assessment was not yet differentiated by sex. The new population assessment model has produced estimates of population and growth rates that, in some cases, differ substantially from the earlier estimates. The basic models for both growth and recruitment have remained unchanged however. In the sections below, the growth and recruitment models are updated from Clark and Hare (2002). For a more complete treatment the source publication should be consulted. The section on maturity is an update of a similar analysis by Parma (1998).

Recruitment

Although IPHC estimates biomass and sets catch limits for each regulatory area separately, we regard the halibut of the northeast Pacific as a single biological stock because there is no evidence of genetic differences and variations in year-class strength are highly correlated among areas. For those reasons we calculate estimates of total spawning biomass and total recruitment to investigate recruitment dynamics. As explained above, our historical abundance estimates are limited to Areas 2B, 2C, and 3A, so the analysis is perforce limited to these areas.

Halibut recruitment has alternated between high and low “regimes” of productivity over at least the past 70 years. Transitions between regimes most recently occurred in 1947 (from high to low) and 1977 (from low to high). Recent research has linked these productivity regimes to an interdecadal mode of pan-Pacific climate variability termed the Pacific Decadal Oscillation (PDO, Mantua et al. 1997).

A plot of age-six recruits (on a log scale, combining the three core areas) and the annual PDO index is shown in Figure 22. A plot of total (age 10+) biomass and age-six recruitment shows little relationship between the two (Fig. 23) and a simple Ricker stock recruitment (S-R)

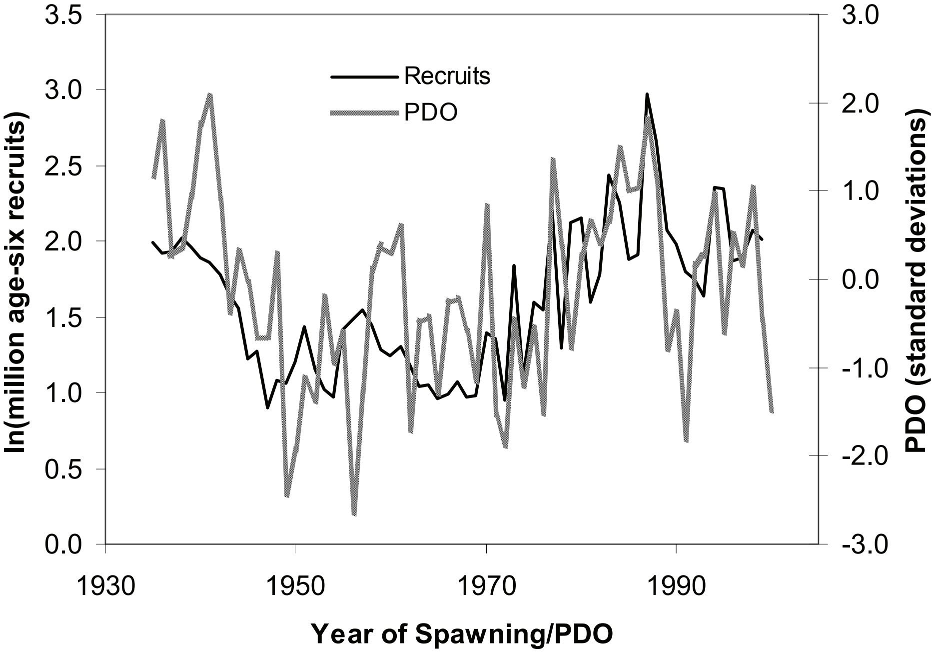


Figure 22. Trend in the annual index of the Pacific Decadal Oscillation (PDO) and the logarithm of age-six recruits for IPHC areas 2B, 2C, and 3A combined.

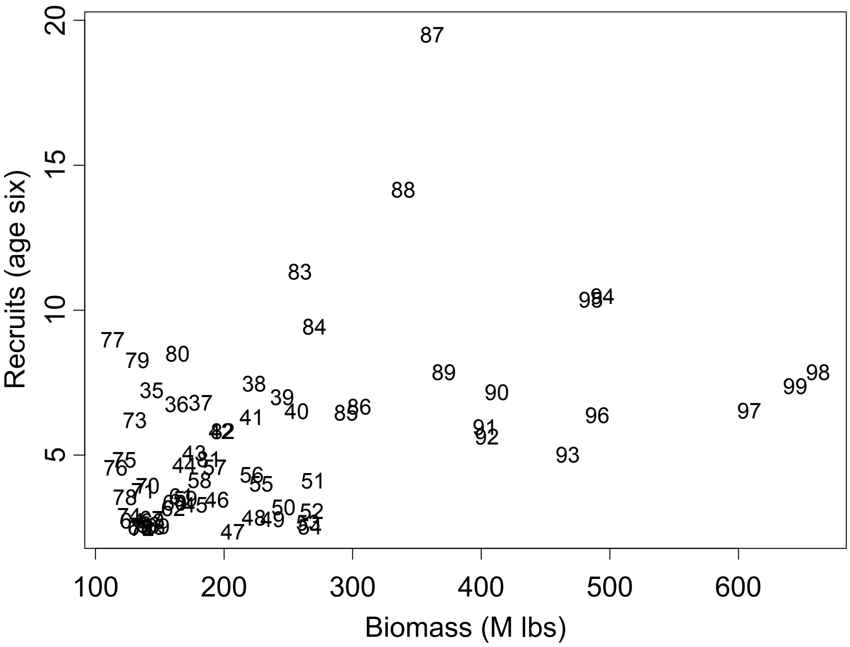


Figure 23. A biomass-recruit plot for halibut for Areas 2B, 2C and 3A combined. The digit numbers refer to the year class during the 20th century.

model fits the data very poorly. Clark and Hare (2002) fitted a series of both density-dependent and density-independent models to halibut recruitment data from different periods. The two basic models were as follows, with *i* indexing period:

$$(1) \qquad \ln(R) = \ln(\alpha_i) - \beta \cdot S + \gamma \cdot PDO$$

$$(2) \qquad \ln(R) = \delta_i + \rho \cdot PDO$$

Model 1 is the usual Ricker S-R model but incorporates regime-specific intercepts and includes the annual PDO index in the year of spawning as a covariate. (Leading and lagging PDO values were found to be not significant.) In this update, the measure of biomass that is used is the total biomass of age 10+ halibut. Model 2 predicts recruitment solely on the basis of regime levels and the annual PDO index with no information on biomass.

Both model fits are greatly improved by allowing regime-specific intercept parameters. In addition to the regime shifts in 1947 and 1977, additional intercept parameters for both models were allowed in 1958 and 1970. The reasoning is that this was a period of unreliable bycatch estimates and it is quite possible that our corrections for bycatch (which are added to the number of age-six recruits) do not capture the full effect. The fit of the two models is very similar (Fig. 24). Model fitting statistics—lower values of the Akaike Information Criterion (AIC) and Schwarz' Bayesian Criterion (SBC)—favor Model 2:

Model 1				Model 2			
Term	Years	Value	SD	Term	Years	Value	SD
$\ln(\alpha_1)$	1935-1946	-3.01	0.13	δ_1	1935-1946	1.67	0.08
$\ln(\alpha_2)$	1947-1958	-3.38	0.16	δ_2	1947-1958	1.37	0.07
$\ln(\alpha_3)$	1959-1970	-3.38	0.13	δ_3	1959-1970	1.17	0.06
$\ln(\alpha_4)$	1971-1976	-2.91	0.13	δ_4	1971-1976	1.55	0.10
$\ln(\alpha_5)$	1977-1998	-2.69	0.18	δ_5	1977-1998	2.00	0.05
β	1935-1998	0.003	0.0004	No β			
γ	1935-1998	0.127	0.03	ρ	1935-1998	0.128	0.04
AIC		-70.7		AIC		-72.2	
SBC		-55.7		SBC		-59.3	

Recruitment for halibut is clearly driven by environmental conditions. The best fit to the recruitment data uses no information on spawning stock size. The best fit for a model using spawning stock as a predictor has regime specific parameter values further verifying the importance of the PDO to halibut recruitment. For the population dynamics simulations described later, recruitment is modeled as a purely environmentally driven process.

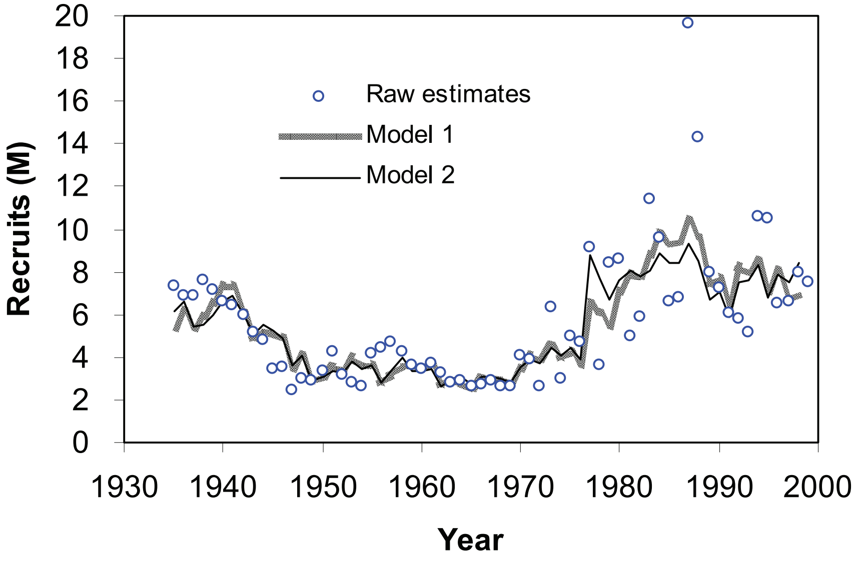


Figure 24. Raw halibut recruitment estimates and values predicted by fits of model 1 and model 2. See text for details.

Growth

Long-term changes in size at age have long been noted for halibut. Halibut of both sexes and all ages 8 and older are substantially smaller than halibut of the same sex and age 30 years ago. However, halibut of the same size at age were seen in the 1920s and 1930s. Clark and Hare (2002) estimated trends in growth using a simple linear model of growth with time varying parameters. Separate models were fitted for IPHC Areas 2B and 3A. The models were of the form:

$$w_a = w8_y + \sum_{9}^{a} GI_y$$

in which weight at age w_a is modeled as the sum of mean weight at age 8 ($w8_y$) and annual growth increments (GI_y) thereafter. Mean weight at age 8 and growth increments were estimated every 10 years between 1920 and 2000 and annual values were then interpolated. The resulting time series for both growth parameters were then plotted against environmental and stock indices. Both growth parameters showed the strongest linear relationship with total numbers of adult halibut (age 10+) with little evidence for an environmental influence. Figure 25 illustrates growth trends for males and females in IPHC Areas 2B and 3A and the contemporary trajectories of total adult numbers. A slightly modified form of this relationship between population size and growth increment is used to model growth dynamics in the simulation modeling described below.

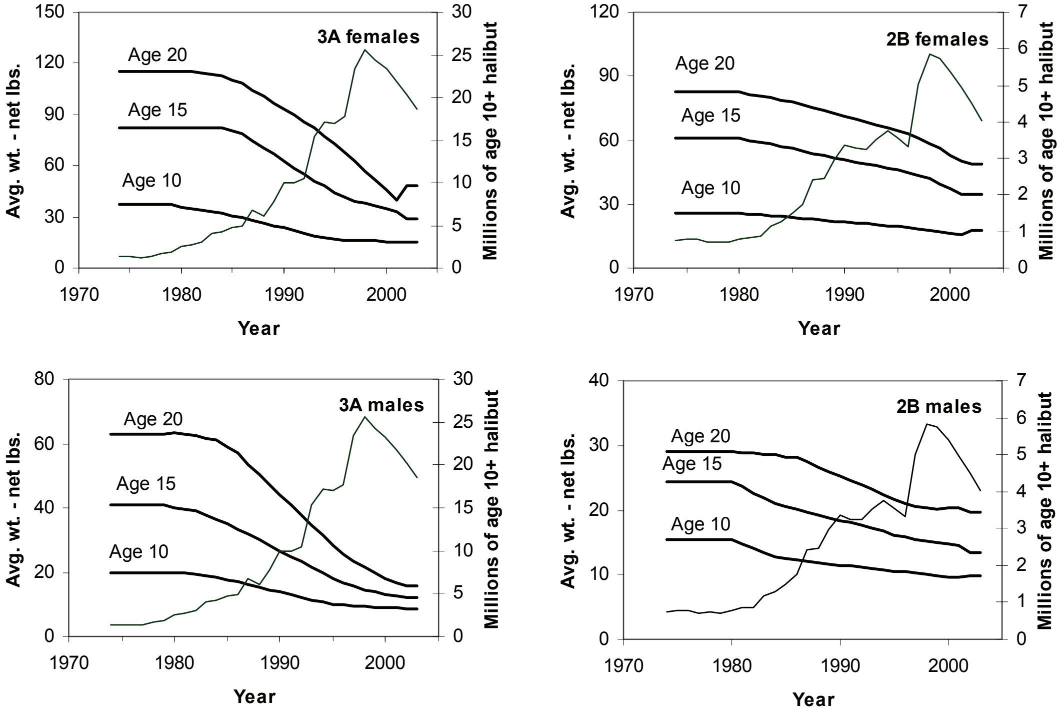


Figure 25. Halibut weight at age from 1974 to 2004. Trend lines are shown for three ages and both sexes for IPHC areas 2B and 3A. The contemporary trajectory of total numbers of adult halibut (ages 10+) is shown as a thin line.

Maturity

Halibut maturity at age has been examined several times, most recently by Parma (1998). All previous analyses were conducted using halibut ages determined by surface readings. We re-examined maturity at age in Areas 2B and 3A during several time periods to look for time trends. For comparison we also examined maturity at length. The available maturity data were grouped as follows: 1963-1966, 1976-1983, 1992-1996, 1997-2001, 2002-2003. To estimate proportion mature p at age/length, logistic functions were fitted to the data for each period and region. The form of the function was as follows:

$$p = \frac{1}{1+\exp\left(-k \cdot \left(A - A_{50}\right)\right)} \qquad \text{or} \qquad p = \frac{1}{1+\exp\left(-k \cdot \left(L - L_{50}\right)\right)}$$

where k is a slope parameter and A_{50} is the age, and L_{50} is the length, at which 50% of the females are mature. To avoid mixing ages from different reading techniques, only surface ages were used for the periods prior to 2001 and only break and burn ages were used for 2002-2003.

The results are illustrated in Figure 26 and maturity function parameter estimates are given in Table 3. Despite the differences in time periods and aging techniques, female maturity at age has been remarkably consistent over time and between Areas 2B and 3A. These time periods capture the extreme in changing growth rates over time and show that maturity is likely determined mainly by age and not by size. The age at which 50% of females attain maturity varies from a low of 10.47 in 2B in the early 90's to a high of 12.27 in 2B in 2002-2003. In Area 2B, the higher age at 50% maturity might be attributed to the new aging technique but a similar increase in A_{50} was not found in Area 3A. The difference in assigned ages is not appreciable until after approximately age 15 and this accounts for why there is no systematic difference between the different aging types. For the harvest rate simulations, we use a single maturity schedule with the parameter estimates from the logistic model fitted to all 2002-2003 data for both areas (A_{50} = 11.59, k = 0.563).

Table 3. Time periods, sample sizes, and parameter estimates for logistic function fits to Pacific halibut maturity data at age (A) and length (L).

Area	Years	n	A_{50}	k	L_{50}	k
2B	1963-1966	647	11.96	0.515	119.71	.111
2B	1976-1983	753	10.99	0.772	111.14	.100
2B	1992-1996	3581	10.47	0.674	97.60	.093
2B	1997-2001	5419	10.78	0.583	93.65	.099
2B	2002-2003	2124	12.27	0.555	101.36	.107
2B	All years	12528	10.97	0.592	98.15	.089
3A	1963-1966	2538	10.45	1.043	119.59	.169
3A	1976-1983	3514	11.62	0.887	125.98	.129
3A	1992-1996	4389	10.91	1.002	92.09	.122
3A	1997-2001	5508	10.66	0.789	85.44	.095
3A	2002-2003	2222	10.83	0.527	85.25	.091
3A	All years	18175	10.93	0.822	96.53	.060
						.
All areas	2002-2003	4347	11.59	0.563	97.63	.070
All areas	All years	30704	10.91	0.711	93.37	.093

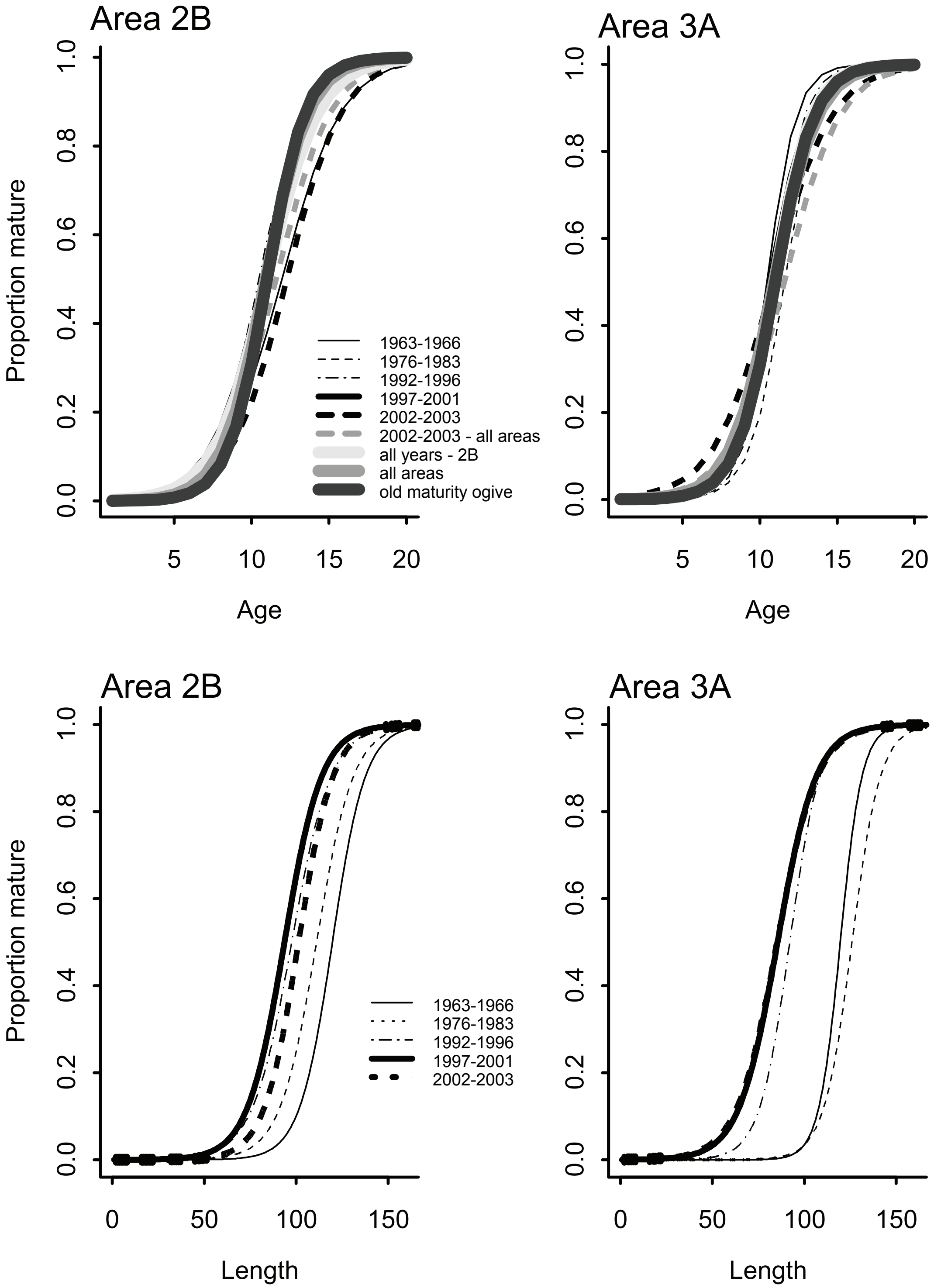


Figure 26. Maturity at age and maturity at length ogives for Pacific halibut in Areas 2B and 3A for a variety of time periods.

There has been a highly significant change in maturity at length. Length at 50% maturity ranged from a high of 120 cm in the 1960s to a low of 94 cm in the late 1990s in Area 2B and from a high of 126 cm in the 1970s to a low of 85 cm in the late 1990s in Area 3A. As shown by the 2002-2003 break and burn data, the declining trend in maturity at length has stopped and even rebounded slightly in Area 2B. Based on these results, maturity is modeled as an age-dependent process in the simulation modeling conducted for the harvest policy analysis.

Harvest policy

Since 1985, the IPHC has followed a constant harvest rate (CHR) policy to determine annual available yield, termed the Constant Exploitation Yield (CEY). The harvest rate, which is the fraction of the exploitable biomass allowed to be harvested annually, has changed over time, from 0.35 in 1985 to 0.30 in 1993 to 0.20 in 1996 to a provisional rate of 0.25 in 2003 to the current value of 0.225 set in 2004. Prior to the CHR policy, harvests were set as a percentage of the estimated annual surplus production (ASP). The ASP policy was implemented at a time of historically low biomasses (Deriso and Quinn 1985). The change to a CHR policy occurred once the stocks were considered to have been rebuilt.

A constant harvest rate policy has a number of attractive features. The CEY rises and falls smoothly with the biomass; catches are automatically scaled down at lower biomasses and increased during periods of high biomass levels. Yields near the theoretical maximum sustainable yield can be taken across a broad range of harvest rates. In a number of simulation studies, a CHR policy has been shown to be quite robust to climate induced variability in productivity of the stock (Walters and Parma 1996, Hilborn and Walters 1992). A CHR policy has also been well received by the industry – it is relatively simple to understand and the halibut fishery has enjoyed a sustained period of high yields.

Between 2002 and 2004, the IPHC staff developed a modified CHR policy, termed the Conditional Constant Catch (CCC) harvest policy (Clark and Hare 2004). The CCC harvest policy was developed to provide more stable catch quotas than the CHR policy used by the IPHC for the past 20 years. The defining features of the CCC policy were an upper cap on quotas at high biomass levels and minimum biomass limits and thresholds at low biomass levels (Hare and Clark 2003). Ultimately, the policy was rejected for use in the halibut fishery. Much of the simulation work done during analyses of the CCC policy is relevant to the current CHR policy and is summarized in this report.

In the years since a CHR policy was adopted, numerous changes have taken place in the halibut stock assessment, and in our understanding of the population dynamics of the stock. With each change, the harvest policy is re-evaluated and an appropriate harvest rate determined. Factors such as density dependent growth and recruitment regimes present novel challenges in modeling halibut population dynamics. The harvest policy builds upon an understanding of the long term dynamics of the stock, and is investigated using a simulation model that incorporates time varying stock dynamics. For this reason, the harvest policy is developed based on the productivity of the IPHC “core” areas, i.e., Areas 2B, 2C and 3A.

Implementation of a minimum biomass threshold and limit

The IPHC considers first and foremost the impact of the harvest policy on female spawning biomass. The approach taken is one of avoidance of dropping below the minimum observed historical level. This is different from the philosophy where harvest control rules are based on a more theoretical construct: spawning biomass per recruit. Within the three areas being analyzed, halibut populations rebounded from the minimum spawning biomasses of the early 1970s to the high levels observed for the past 15-20 years. We can have some confidence therefore of stock dynamics at those spawning biomass levels, but not at lower levels. There is no compelling reason to allow spawning biomass to drop below the minimum limit.

In keeping with the global movement towards precautionary management, an additional biomass safeguard was investigated and adopted. The terms "threshold" and "limit" have come into use in fisheries management to define levels at which extra conservation measures are implemented. There is no universally accepted definition for the terms and they are often used interchangeably. For the purposes of the Pacific halibut harvest policy, we define threshold as a level at which more conservative harvest rates begin to apply, and limit as a biomass level at which all fishing on the stock ceases.

There are at least two rationales for establishing reasonable minimum biomass safeguards. A fairly common threshold is B_{MSY}, i.e., the equilibrium biomass when fishing at the MSY rate. A common limit associated with this threshold is 0.5 B_{MSY}. This is somewhat problematic for halibut due to its alternating productivity regimes as well as density dependent growth. A second rationale for selecting a limit and threshold has to do with what has historically been observed for the stock. If a stock has been monitored long enough to observe a descent to, and recovery from, a low point then that low point may be a "safe" minimum limit. We followed this second rationale in establishing a minimum biomass threshold and limit for Pacific halibut.

The minimum observed spawning biomasses for the three IPHC core areas all occurred in the mid 1970s, approximately 9 million pounds in 2B, 13 million pounds in 2C and 42 million pounds in 3A. By definition, these become the spawning biomass limits. These are common sense limits. In the IPHC harvest policy, the target harvest rate is linearly scaled downwards once spawning biomass reaches the threshold. In simulations, this was found to be very effective in returning the spawning biomass to at least the threshold in a short time without greatly affecting yield. We tested several thresholds, ranging from 1.25 to 2.00 times the limit. A threshold equal to 1.5 times the limit performed well in simulations, producing lower variability in yield than higher or lower values (Hare and Clark 2003). The IPHC modified CHR policy is illustrated in Figure 27.

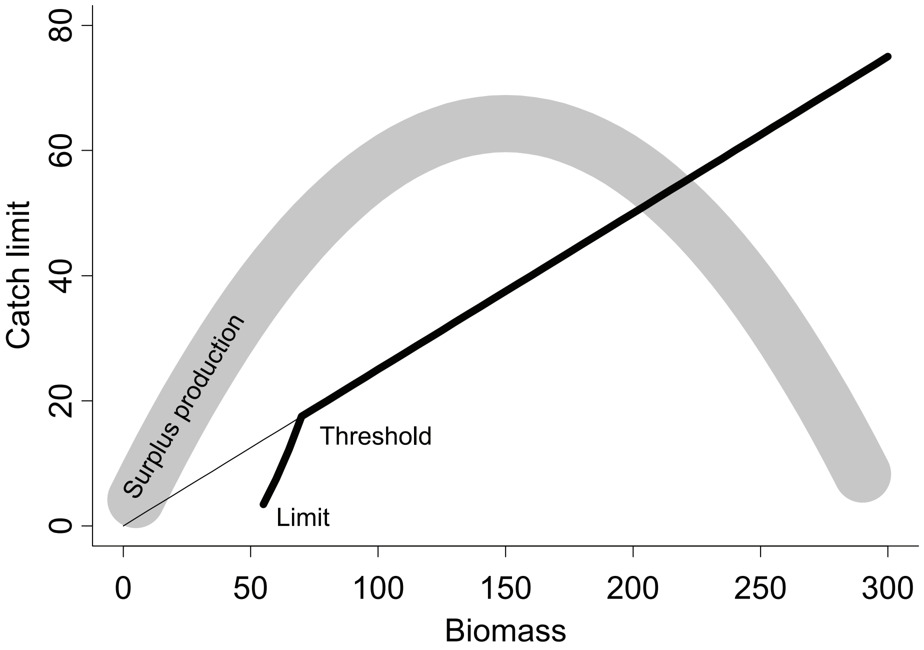


Figure 27. Illustration of how a modified constant harvest rate policy would operate. Above the threshold the harvest rate is equal to the slope of the angled line. The harvest rate scales down to a rate of 0.0 as the biomass drops below the threshold and approaches the limit. Theoretical relaionship between biomass and surplsus production is illustrated in the background. All units are arbitrary.

Performance of harvest policy

The goal of the halibut harvest policy is to achieve a high level of yield while at all times maintaining a healthy female spawning biomass (all subsequent references to spawning biomass imply female spawning biomass). Over the past few years there have been several advances in our understanding of halibut population dynamics. Several substantive changes have also occurred in the stock assessment model used to estimate population. Among the most important changes since the last published analysis of the harvest policy (Sullivan et al. 1997) are: a lower natural mortality rate, independent accounting of sexes, quantification of aging error, length-specific selectivity, and the new views about factors affecting growth and recruitment. A constant harvest rate policy has served the halibut population well but needs to be re-examined in light of these changes.

Simulation model

The harvest policy is investigated via standard population dynamics simulation. The simulation model has separate accounting of males and females and sex-specific growth rates. Stock dynamics are modeled as described above, with recruitment controlled by environmental conditions and growth varying in a density-dependent manner. Separate simulation models are run for each area.

Under the modern assessment model, average recruitment (in millions of age-6 recruits) for the periods before and after the regime shift of 1977 are as listed in the following table. Recruitment estimates have been adjusted for bycatch losses to the commercial groundfish fisheries.

Area	1968-1976	1977-1998	All years
2B	1.18	2.35	2.02
2C	1.17	2.16	1.89
3A	2.80	7.39	6.06
Combined	5.14	11.84	9.98

Area 3A accounts for approximately 60% of total recruitment to the three regions while Areas 2B and 2C each account for about 20%; annually simulated recruitment was divided among the three areas in these proportions. We simulated the duration of a PDO-associated climate regime by drawing from a uniform (15, 30) distribution. For the harvest rate simulations, we modeled recruitment as alternating regimes of high and low productivity. During productive regimes average recruitment into the three regions was 11.84 million age-6 halibut (50% male and 50% female), and during unproductive regimes average recruitment was 5.14 million age-6 halibut. Variability within regimes was generated using the following relationships:

$$R_6 = \exp(\ln(\mu_i)) + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + e_t$$

where μ is average recruitment, *i* indexes regime, lag-1 autocorrelation ρ = 0.1 (estimated from the observed recruitments), ε_t is autocorrelated error, e_t is normal process error with σ_ε =0.4 (ρ and σ_ε obtained from residuals of recruitment Model 2 described above).

The method devised by Clark and Hare (2002) to estimate density dependent effects on growth was altered slightly to account for changes in size at age since that analysis. The growth model developed for this analysis contains three growth parameters: mean size at age 6, an annual growth increment between ages 6 and 20, and an annual growth increment between ages 21 and 30. Growth in weight is therefore linear from age 6 to 20 at one rate and then linear at a different (and lower) rate from ages 21 to 30. There are separate parameter estimates for each area (Area 2B, 2C, 3A and all three regions combined), sex and fishery (survey and commercial). For each area, sex and fishery type, there are high and low values for the two annual growth increments and these are related to the number of adult animals in the stock. Size at age 6 varies by sex, region and fishery but is constant across population size. The parameter estimates, along with minimum and maximum mean weight at age for ages 20 and 30 are given in Table 4. Weight at age from the survey is used to represent the weight of halibut in the population (e.g., for computing spawning biomass and total biomass), weight at age from the commercial fishery is used to represent the weight of halibut in the catch (e.g., for computing exploitable biomass and numbers of fish caught).

Selectivity is assumed to be a fixed function of length. Commercial selectivity assumes the current 81 cm size limit and approximates the selectivities estimated in the modern stock assessment. For areas 2C and 3A the selectivity schedule is as follows:

$$\begin{array}{ll} Sel = 0 & L \le 80 \\ Sel = \dfrac{(L\text{-}80)}{40} & 81 \le L < 120 \\ Sel = 1 & L \ge 120 \end{array}$$

Area 2B has consistently shown higher selectivity at length than Areas 2C and 3A. The schedule used in Area 2B is:

$$\begin{array}{ll} Sel = 0 & L \le 75 \\ Sel = \dfrac{(L\text{-}75)}{35} & 76 \le L < 110 \\ Sel = 1 & L \ge 110 \end{array}$$

A survey selectivity schedule was also used to evaluate the effect of dropping the minimum size limit (and making the assumption that commercial selectivity would then be comparable to survey selectivity). The survey selectivity schedule is similar to the commercial selectivity schedule for Areas 2C and 3A but the lengths at which selectivity equals zero and one are 70 cm and 130 cm, respectively. The same survey selectivity schedule was used for all three areas.

Selectivity at age is computed by combining the fixed selectivity at length schedules with estimated length at age. Length at age (L_a) is computed from mean weight at age (w_a) by inverting the length-weight relationship (Clark 1992a):

$$L_a = \sqrt[\frac{1}{3.24}]{\frac{w_a}{0.00000692}}$$

In general, this method underestimates mean length compared to the usual method of integrating across length using the observed mean length at age and standard deviation. An analysis of survey data showed that the underestimate of mean length computed in this manner is at most 2%; this difference is quite minor and allows us to use our growth model point estimates of mean length.

Table 4. Growth parameters used to establish current and past weight at age by area, sex, fishery and age. See text for details on model.

Area	Sex	Min-Max	Age 10+ fish in population	Age 6 weight (net lbs)	7-20 growth increment (net lbs)	21-30 growth increment (net lbs)	Age 20 wt (net lbs)	Age 30 wt (net lbs)
Setline survey parameters								
2B	F	Min	1	9.0	3.0	2.0	51.0	71.0
	F	Max	6	9.0	5.2	2.5	81.8	106.8
	M	Min	1	6.0	1.0	0.5	20.0	25.0
	M	Max	6	6.0	1.7	1.0	29.8	39.8
2C	F	Min	1	7.0	4.0	1.5	63.0	78.0
	F	Max	6	7.0	6.5	2.5	98.0	123.0
	M	Min	1	5.5	1.3	1.0	23.7	33.7
	M	Max	6	5.5	2.5	1.0	40.5	50.5
3A	F	Min	2	7.0	2.5	2.0	42.0	62.0
	F	Max	25	7.0	8.0	2.0	119.0	139.0
	M	Min	2	7.0	1.0	0.5	19.0	24.0
	M	Max	25	7.0	4.0	2.0	61.0	81.0
All	F	Min	4	8.0	3.0	2.5	50.0	75.0
	F	Max	37	8.0	6.5	2.5	99.0	124.0
	M	Min	4	5.5	1.0	1.0	19.5	29.5
	M	Max	37	5.5	3.0	1.0	47.5	57.5
Commercial fishery parameters								
2B	F	Min	1	15.0	2.3	2.0	47.2	67.2
	F	Max	6	15.0	4.8	2.5	82.2	107.2
	M	Min	1	13.0	0.5	0.6	20.0	26.0
	M	Max	6	13.0	1.2	0.9	29.8	38.8
2C	F	Min	1	13.0	3.0	2.4	55.0	79.0
	F	Max	6	13.0	6.3	2.4	101.2	125.2
	M	Min	1	14.0	0.7	1.0	23.8	33.8
	M	Max	6	14.0	2.0	1.0	42.0	52.0
3A	F	Min	2	13.0	2.0	1.0	41.0	51.0
	F	Max	25	13.0	7.5	4.0	118.0	158.0
	M	Min	2	13.0	0.5	0.5	20.0	25.0
	M	Max	25	13.0	3.5	2.0	62.0	82.0
All	F	Min	4	13.0	2.5	1.5	48.0	63.0
	F	Max	37	13.0	6.3	3.0	101.2	131.2
	M	Min	4	13.0	0.5	0.6	20.0	26.0
	M	Max	37	13.0	2.0	1.3	43.8	56.8

Simulation results

The combination of alternating recruitment regimes and density dependent growth response results in complex population dynamics. To some extent, halibut are cushioned against a rapid decline in population biomass by the higher growth rates achieved at low population sizes. The prime age classes in the catches — ages 11 to 17 — can weigh more than twice as much at a given age at low population numbers than when the population is at a high level. We did not extrapolate beyond the data to estimate growth rates at population numbers smaller than we have observed.

In Clark and Hare (2004), harvest rates up to a maximum of 0.40 were investigated; values higher than 0.40 sometimes drove spawning biomass below the limit, but values equal to or less did not. Thus, a harvest rate of 0.40 functions in the same manner as the “maximum fishing mortality threshold” that is defined under National Standard 1 for NMFS’ managed groundfish stocks. By that definition, harvest rates above the reference value of 0.40 would constitute “overfishing”. By restricting allowable harvest rates to the range of 0.00 to 0.30, allowance is made for observation error in estimates of exploitable biomass. Analysis of retrospective patterns in halibut assessments indicate initial stock biomass estimates have a coefficient of variation of 10-15% (Clark and Hare 2005). Thus, even with a persistent underestimate of the true stock biomass, restriction of harvest rates to a maximum of 0.30 would ensure that the maximum rate of 0.40 would not be reached.

For each harvest rate and area, simulations run forward (from currently estimated numbers and weight at age) for 150 years to establish equilibrium conditions, and performance statistics are tabulated for the next 100 years. Two hundred Monte Carlo replicates are run and results are averaged across replicates. Many population and catch indices are tracked in the simulations; for purposes of selecting a harvest rate, four sets of indicators are used: average catch, frequency of spawning biomass reaching the threshold, realized average harvest rate, and long term average spawning biomass relative to unfished level. Other indicators of interest but not reported here include, e.g., female proportion in the catch, numbers of age 20+ fish remaining in the population, average weight of fish in the catch, etc.

A reference set of simulations and results are developed for the “Most Likely” scenario, i.e., one incorporating all dynamics as outlined above. In addition to reporting results for the “Most Likely” scenario, a second set of results are shown for an alternative scenario—the “Low Growth” scenario. This scenario is utilized to test the robustness of the harvest policy to what is likely the most critical of the dynamic life history traits: density dependent growth. Under this scenario, it is assumed that the current low growth rates—attributed to large numbers of fish in the population—are instead the result of some fundamental ecosystem change. Alternatively, a low growth rate might occur if the halibut population had been “culled” of fish with a genetic disposition towards rapid growth. This alternative scenario is believed to be the most realistic alternative scenario. In previous analysis, other scenarios were examined, including redistributed recruitment among areas and continuous low recruitment levels (Hare and Clark 2003).

The dynamics of the modeled population and effect of the harvest policy are illustrated in Figure 28. The figure shows a single 100-year run for Area 3A at a harvest rate of 0.25. The top panel shows alternating regimes of productivity. The second panel shows how recruitment varies within and between regimes. The third panel illustrates how the mean weight at age for 14 year old halibut change over time in response to varying population numbers. The fourth panel shows yield varying between a low of approximately 25 million pounds to a maximum of nearly 70 millions pounds. Due to their larger size, females form a larger fraction of the yield. The trajectory of spawning biomass is illustrated in the fifth panel. Two lines are shown here – the thin line shows what the biomass trajectory would have been absent the minimum spawning biomass threshold and limit. The thick line show the trajectory when the threshold and limit are imposed and the harvest rate is reduced once spawning biomass reaches the threshold. It is clear that the imposition of the reduced harvest rate acts to limit the downward tendency of the spawning biomass. The bottom panel shows the realized harvest rate, i.e., any excursions below 0.25 represent years that the spawning biomass reached the threshold and triggered a harvest rate reduction.

Performance statistics for both the Most Likely and Low Growth scenarios are summarized in Table 5. Differences between growth scenarios are greatest in Area 3A because the density

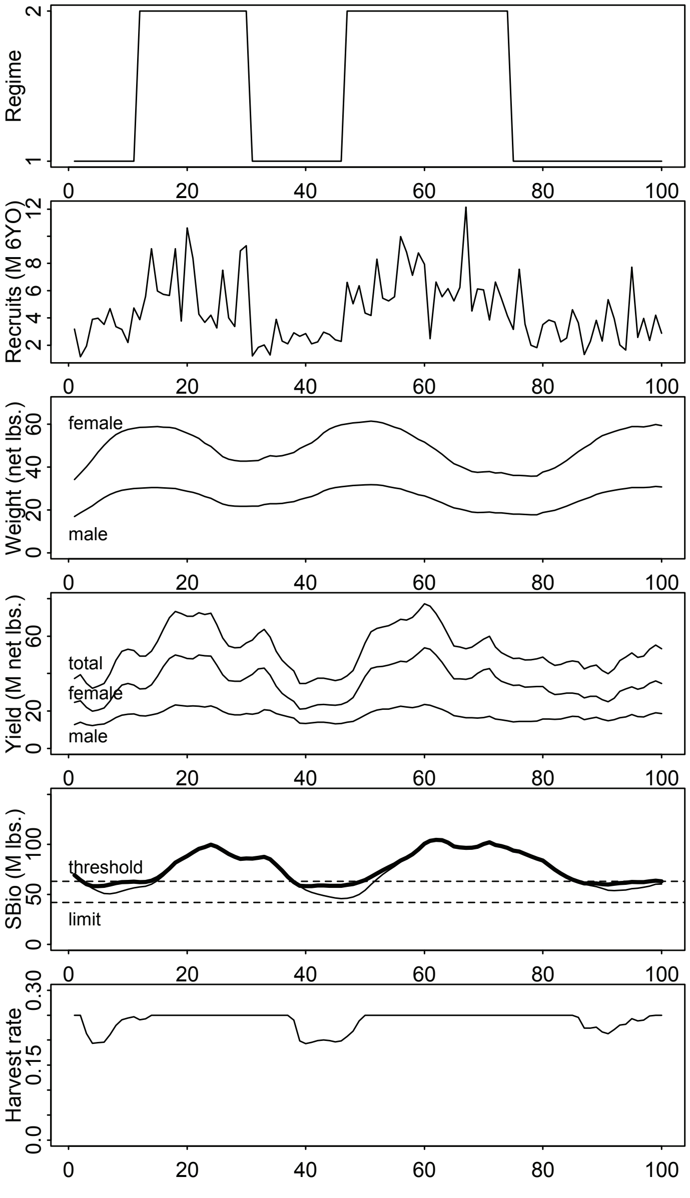


Figure 28. An illustration of the simulations conducted to test the harvest rate policy. The x axis in all plots is years. The top panel plot shows the duration of alternating regimes. Below that is shown total age 6 recruits, next is average weight of a 14 year old fish (male and female), next is yield (male, female and total), next is Spawning Biomass (SBio, females only, threshold and limit biomass reference levels shown as horizontal lines), and bottom panel shows actual harvest rate. These simulations are for Area 3A at a target harvest rate of 0.25.

Table 5. Performance statistics for a range of harvest rates under the Most Likely (Density Dependent (DD) Growth) and Low Growth scenarios.

DD Growth				Low Growth			
	Average annual yield (million pounds)				Average annual yield (million pounds)		
HR	2B	2C	3A	HR	2B	2C	3A
0.000	0	0	0	0.000	0	0	0
0.200	15.3	16.2	45.0	0.200	11.5	11.5	23.5
0.225	16.0	17.1	49.2	0.225	11.9	11.9	24.4
0.250	16.6	17.8	52.9	0.250	12.2	12.2	25.2
0.275	17.1	18.3	56.1	0.275	12.5	12.4	25.8
0.300	17.5	18.8	58.9	0.300	12.7	12.6	26.4

	Average spawning biomass (fraction of HR=0.00 biomass)				Average spawning biomass (fraction of HR=0.00 biomass)		
HR	2B	2C	3A	HR	2B	2C	3A
0.000	105.0	122.6	268.0	0.000	100.4	117.3	250.7
0.200	0.24	0.27	.036	0.200	0.22	0.27	0.36
0.225	0.20	0.24	0.32	0.225	0.20	0.24	0.34
0.250	0.18	0.22	0.30	0.250	0.18	0.22	0.32
0.275	0.16	0.20	0.28	0.275	0.17	0.21	0.30
0.300	0.15	0.18	0.26	0.300	0.15	0.19	0.29

	Average actual harvest rate				Average actual harvest rate		
HR	2B	2C	3A	HR	2B	2C	3A
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.200	0.200	0.200	0.199	0.200	0.199	0.198	0.193
0.225	0.224	0.222	0.222	0.225	0.220	0.218	0.211
0.250	.0244	0.242	0.240	0.250	0.238	0.236	0.228
0.275	0.261	0.258	0.256	0.275	0.253	0.252	0.244
0.300	0.276	0.273	0.270	0.300	0.268	0.267	0.258

dependent variation in growth is greatest there. The results show the expected pattern of increasing catch and decreasing spawning biomass. Average annual yield increases rapidly from a harvest rate of 0.00 to 0.20 and then increases only moderately up a harvest rate of 0.30. Average spawning biomass declines sharply in response to fishing. At a harvest of 0.20, average spawning biomass declines to 24-36% of the unfished average. At a harvest rate of 0.30, average spawning biomass drops as low as 15% of the unfished value in Area 2B. The realized harvest rate begins to drop below the target harvest rate at a target harvest rate of 0.20 and accelerates rapidly thereafter. This is illustrated in Figure 29, which depicts the percentage of time that spawning biomass drops below the minimum biomass threshold. Every time the threshold is reached, the realized harvest rate is less than the target harvest rate. Under the Most Likely scenario, at a target harvest rate of 0.25 the minimum biomass threshold is reached 21-29% of the time in the three IPHC areas. At a harvest rate of 0.30, the threshold is reached approximately twice as often as at a rate of 0.25. However, at a slightly reduced harvest rate of 0.225, the frequency of reaching the threshold is less than half the frequency at 0.25.

Discussion

A CHR strategy has been used to establish catch limits at the IPHC for 20 years. The harvest rate has ranged from a low of 0.20 to a high of 0.35. The latter value, based on old estimates of recruitment and growth, appears now to be clearly too large. The optimal value thus appears to be within a range of 0.20-0.30. On the basis of the relative infrequency of reaching the minimum biomass threshold and the relatively minor difference in average yield between a harvest rate

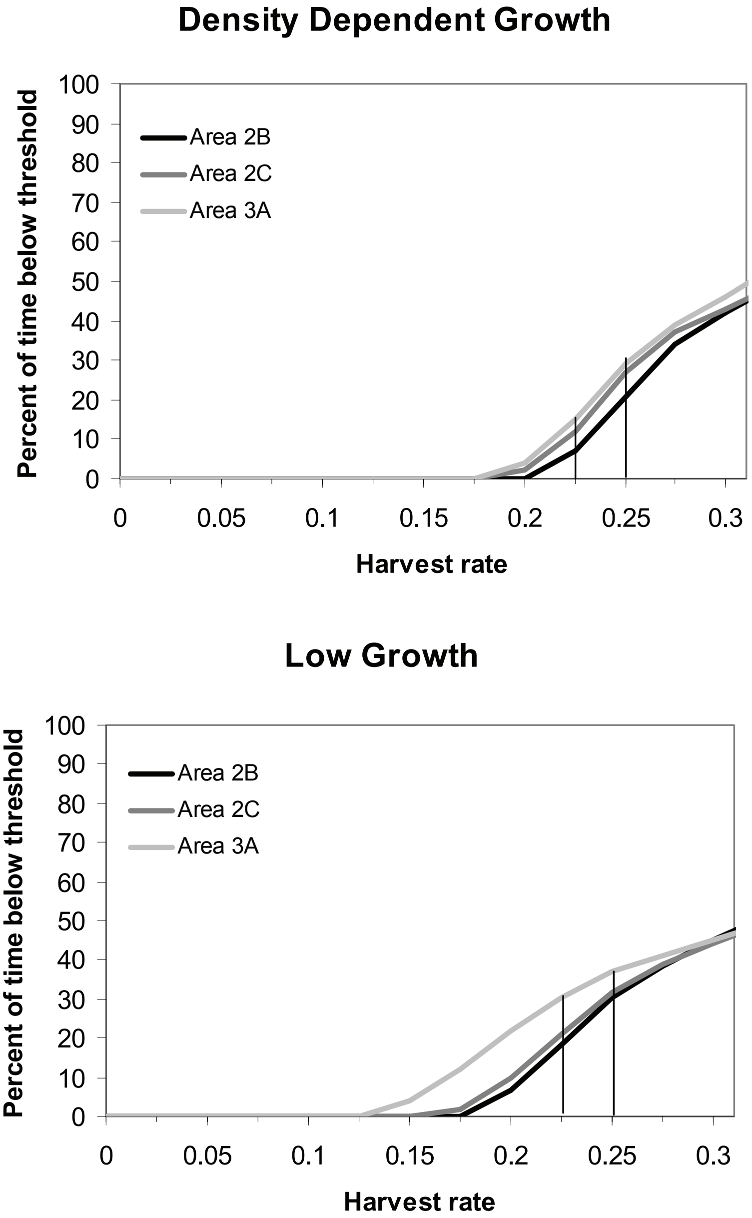


Figure 29. In the harvest rate simulations, the percent of years in which spawning biomass dropped below the minimum spawning biomass across a range of harvest rates. The upper panel illustrates the Most Likely (Density Dependent Growth) scenario, the bottom panel shows the Low Growth scenario.

of 0.225 and 0.25, a target harvest rate of 0.225 appears appropriate for the halibut stock in the core areas. Under the Low Growth scenario, there is an increased frequency of reaching the minimum biomass threshold, particularly in area 3A. However the expected occurrence is no greater than 30% of the time. Given the precautionary application of the reduced harvest rate triggered when the threshold is reached, this appears to be a tolerable level of risk.

Support for a harvest closer to the lower end of the range is provided from a couple of other perspectives. At harvest rates over 0.225, the average spawning biomass in Areas 2B and 2C drops below 20% of the unfished level. A number of published studies have suggested that for groundfish, average spawning biomass should remain in the range of 20-60% (e.g., Clark 1991). Secondly, yield per recruit and spawning biomass per recruit analyses were conducted using different sets of growth rates (Hare and Clark 2005b). Based on those analyses, a harvest rate in the range of 0.15-0.20 would be recommended as spawning biomass per recruit was reduced to 35-40% of the unfished level.

The target harvest rate of 0.225 applies to the IPHC core areas of 2B, 2C and 3A. In Area 2A, where biomass is leveraged off the 2B estimate, the target harvest rate is also set 0.225. For Areas 3B and 4, a lower target harvest rate of 0.200 is currently used. Under the precautionary principle, a more conservative fishing rate should be applied since far less is known about the productivity potential of those areas. There is not currently enough historical data to conduct a dynamic harvest policy analysis for areas 3B and 4. In the absence of a dynamic analysis, a reasonable fallback is the spawning biomass per recruit equilibrium analysis using weight at age data from the appropriate regions. Since size at age in Areas 3B and 4 is not greatly different from Area 3A the same basic results would apply – a harvest rate between 0.15 and 0.20 is appropriate.

In this harvest rate analysis, we have attempted to capture the dynamics of the halibut stock in establishing an appropriate harvest rate. Numerous sources of uncertainty in addition to the “Most Likely” scenario were explored and reported upon here and elsewhere (Hare and Clark 2001, 2003). These included uncertainties in density dependent growth response, future levels of recruitment as well as distribution of recruitment among areas, stock- and environment-recruitment relationships, selectivity curves, etc. There are other forms of uncertainty that could be considered. These include both biological and operational factors. The biological uncertainty is likely to have been mostly captured in the simulations. However, uncertainty in the annual stock assessment—other than putting observation error on biomass estimates—is not incorporated. Neither is management uncertainty, e.g., not faithfully setting catch limits at the computed levels. Testing the robustness of the harvest policy to these types of uncertainty is the goal of the developing paradigm of “management strategy evaluation” (Punt and Smith 1999, Cooke 1999). Exploring the IPHC harvest policy in such a framework is a logical next step.

References

Bell, F. H. 1981. The Pacific halibut. Alaska Northwest Publ. Co., Anchorage, Alaska.

Blood, C. L. 2005. 2004 sport fishery. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004:43-50.

Chapman, D. G., Myhre, R. J., and Southward, G. M. 1962. Utilization of Pacific halibut stocks: estimation of maximum sustainable yield, 1960. Int. Pac. Halibut Comm. Rep. 31.

Clark, W. G. 1992a. Validation of the IPHC length-weight relationship for halibut. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 1991:113-116.

Clark, W. G. 1992b. Estimation of halibut body size from otolith size. Int. Pac. Halibut Comm. Sci. Rep. No. 75.

Clark, W. G. 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. Can. J. Fish. Aquat. Sci. 56:1721-1731.

Clark, W. G. 2000. Estimates of reductions in halibut recruitment due to bycatch. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 1999: 139-147

Clark, W. G. 2002a. Comparison of snap-hook and fixed-hook CPUE. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2001:191-198.

Clark, W. G. 2002b. Comparison of setline survey CPUE in standard survey subareaas and entire regulatory areas. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2001:379-384.

Clark, W. G. 2003. A model for the world: 80 years of model development and application at the International Pacific Halibut Commission. Natural Resource Modeling 16:491-503.

Clark, W. G. 2005. Estimates of length-specific commercial selectivity from historical marking experiments. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004:91-102.

Clark, W. G. 2006a. Overview of the IPHC catch sampling strategy. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2005:83-84.

Clark, W. G. 2006b. Effects of gear type, hook spacing, and hook size on commercial selectivity and catchability. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2005:93-104.

Clark, W. G. 2006c. Analysis of PIT tag recoveries through 2005. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2005:123-134.

Clark, W. G., and Chen, D.G. 2005. Preliminary estimates based on 2004 PIT tag recoveries. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004:199-212.

Clark, W. G., and Hare, S.R. 1998. Accounting for bycatch in management of the Pacific halibut fishery. N. Am. J. Fish. Manage. 18:809-821.

Clark, W. G., and Hare, S. R. 2002. Effects of climate and stock size on recruitment and growth of Pacific halibut. N. Am. J. Fish. Manage. 22:852-862.

Clark, W. G. and Hare, S. R. 2004. A conditional constant catch policy for managing the Pacific halibut fishery. N. Am. J. Fish. Mgmt. 24:106-113.

Clark, W. G. and Hare, S. R. 2005. Assessment of the Pacific halibut stock at the end of 2004. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004: 103-123.

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*). Can. J. Fish Aquat. Sci. 56: 242-252.

Clark, W. G., and Parma, A. M.. 1999. Assessment of the Pacific halibut stock in 1998. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 1998: 89-112.

Clark, W. G., Vienneau, B.V., Blood, C. L., and Forsberg, J. E. 2000. A review of IPHC catch sampling for age and size composition from 1935 through 1999, including estimates for the years 1963-1990. IPHC Tech. Rep. 42.

Cooke, J. G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES J. Mar. Sci. 56: 797-810

Deriso, R. B. and Quinn II, T. J. 1985. Methods of population assessment of Pacific halibut. Int. Pacific Halibut Comm. Scientific Report No. 72, 52 p.

Deriso, R. B. Quinn, T. J., and Neal, P. R. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42:815:824.

Forsberg, J. E. 2001. Aging manual for Pacific halibut: procedures and methods used at the International Pacific Halibut Commission. Int. Pac. Halibut Comm. Tech. Rep. 46.

Fournier, D. A., Sibert, J. R., Majkowski, J., and Hampton, J. 1990. MULTIFAN, a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data fro southern bluefin tuna (*Thunnus maccoyi*). Can. J. Fish. Aquat. Sci. 47:301-317.

Grant, W. S., Teel, D. J., Kobayashi, T., and Schmitt, C. 1984. Biochemical population genetics of Pacific halibut (*Hippoglossus stenolepis*) and comparison with Atlantic halibut (*H. hippoglossus*). Can. J. Fish. Aquat. Sci. 41:1083-1088.

Hamley, J. M., and Skud, B. E. 1978. Factors affecting longline catch and effort. II. Hookspacing. Int.Pac. Halibut Comm. Sci. Rep. 64:15-24.

Hare, S. R. 2001. Updated historical estimates of Pacific halibut biomass and recruitment trends, 1935-1973. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2000: 119-152.

Hare, S. R. and Clark, W. G. 2001. Evaluation of alternative harvest rates. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2000: 153-167.

Hare, S. R. and Clark, W. G. 2003. Issues and tradeoffs in the implementation of a conditional constant catch harvest policy. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2002: 121-161.

Hare, S. R., and Clark, W. G. 2005. Yield per recruit analysis for a sex specific halibut model. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004:171-184.

Hare, S. R., Williams, G. H., and Chen, D. 2004. Bycatch mortality and size distribution extrapolation methodology. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2003: 255-272.

Hauser, L., Spies, I., and Loher, T. 2006. Microsatellite screening in Pacific halibut (*Hippoglossus stenolepis*) and a preliminary examination of population structure based on observed DNA variation. Int. Pac. Halibut Comm. Sci. Rep. 81.

Hilborn, R. and C. J. Walters. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, & Uncertainty. Chapman and Hall, New York.

International Pacific Halibut Commission (IPHC). 1998. The Pacific halibut: biology, fishery, and management. Int. Pac. Halibut Comm. Tech. Rep. 40.

IPHC Staff. 1960. Utilization of Pacific halibut stocks: yield per recruitment. Int. Pac. Halibut Comm. Rep. 28.

Lester, N. P., Shuter, B. J., and Abrams, P. A. 2004. Interpreting the von Bertalanffy model of somatic growth in fishes: the cost of reproduction. Proc. R. Soc. Lond. B 271:1625-1631.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin American Meteorological Society 78:1069-1079.

Maunder, M. N., and Watters, G. M. 2003. A-SCALA: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the eastern Pacific Ocean. Bull. I-ATTC 22:435-582.

Myhre, R. J. 1967. Mortality estimates from tagging experiments on Pacific halibut. Int. Pac. Halibut Comm. Rep. 42.

Myhre, R. J. 1974. Minimum size and optimum age of entry for Pacific halibut. Int. Pac. Halibut Comm. Sci. Rep. 55.

Myhre, R. J., and Quinn, T. J. 1984. Comparison of efficiency of snap gear to fixed-hook setline gear for catching Pacific halibut. Int. Pac. Halibut Comm. Sci. Rep. 69.

Myhre, R. J., Peltonen, G. J., St-Pierre, G., Skud, B. E., and Walden, R. E. 1977. The Pacific halibut fishery: Catch, effort and CPUE, 1929-1975. Int. Pac. Halibut Comm. Tech. Rep. 14.

Parma, A. M. 1998. Re-evaluation of the 32-inch commercial size limit. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 1997: 167-202.

Punt, A. E., and D. S. Butterworth. 1993. Variance estimates for fisheries assessment: their importance and how best to evaluate them. p. 145-162. *In* S.J. Smith, J.J. Hunt, and D. Rivard [ed.] Risk evaluation and biological reference points for fisheries management. Can. Spec. Publ. Fish. Aquat. Sci. 120.

Punt, A. E and Smith, A. D. M. 1999. Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). ICES J. Mar. Sci. 56: 860-875.

Schmitt, C. C., and Skud, B. E. 1978. Relation of fecundity to long-term changes in grwoth, abundance, and recruitment. Int. Pac. Halibut Comm. Sci. Rep. 66.

St-Pierre, G. 1984. Spawning locations and season for Pacific halibut. Int. Pac. Halibut Comm. Sci. Rep. 70.

St-Pierre, G. 1989. Recent studies of Pacific halibut postlarvae in the Gulf of Alaska and eastern Bering Sea. Int. Pac. Halibut Comm. Sci. Rep. 73.

Sullivan, P. J., Parma, A. M., and Clark, W. G. 1999. The Pacific halibut assessment of 1997. Int. Pac. Halibut Comm. Sci. Rep. 79.

Thompson, W. F., and Bell, F. H. 1934. Biological statistics of the Pacific halibut fishery. (2) Effects of changes in intensity upon total yield and yield per unit of gear. Int. Fisheries Comm. Rep. 8.

Thompson, W. F., and Freeman, N. L. 1930. History of the Pacific halibut fishery. Int. Fisheries Comm. Rep. 5.

Trumble, R. J., McGregor, R. I., St-Pierre, G., McCaughran, D. A., and Hoag, S. H. 1990. Sixty years of tagging Pacific halibut: a case study. Pages 831-840 *in* N.C. Parker and five co-editors. Fish-marking techniques. American Fisheries Society, Symposium 7, Bethesda, Maryland.

Walters, C. J. and Parma, A. 1996. Fixed exploitation rate strategies for copnig with the effect of climate change. Can. J. Fish. Aquat. Sci. 53: 148-158.

Williams, G. H. 2005a. Incidental catch and mortality of Pacific halibut, 1962-2004. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004:213-223.

Williams, G. H. 2005b. Revised estimates of the personal use harvest, including new estimates for the subsistence fishery off Alaska. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2004:55-60.

Williams, G. H., Schmitt, C. C., Hoag, S. H., and Berger, J. D. 1989. Incidental catch and mortality of Pacific halibut, 1962-1986. Int. Pac. Halibut Comm. Tech. Rep. 23.

Appendix A. Catalogue of IPHC setline survey data through 1999

Abstract

This paper identifies all sets in the IPHC survey database that were part of systematic (grid) surveys through 1999. It also describes some of the query parameters that often enter into retrievals of grid survey data.

List of early survey stations (1963-1986)

IPHC has conducted a number of systematic setline surveys in various parts of the Commission area in the years 1963-66, 1976-86, and 1993-present. Before 1993 stations were laid out on transects. Since then they have been laid out on a grid, first in triangular clusters and since 1998 as single stations on a regular 10 nmi square pattern. Surveys through 1979 are described by Hoag et al. (1980); evolution of survey design up to the present by Randolph (1998).

The data from survey trips reside in a setline database that contains data from all charters, so retrieving the survey data requires a properly qualified query. Since 1993 all standard survey data have been assigned a code of "SG" (Standard Grid) in the "purpose" field of each station record. For earlier years it is necessary to retrieve the data by vessel, year, and in some cases station number (where a vessel did some survey stations and some other work on the same trip).

Before 1993, offshore survey stations were spaced along parallel transects that ran east-west in Area 2 and north-south in Area 3. Transects were numbered from 1 to 95 in a counterclockwise direction. Stations along a transect were labeled by letter A, B, C, … from inshore to the edge of the shelf. For these stations the "station number" (called "stnno") that appears in the database is actually the transect number. The "station position" (called "stnpos") in the database is the letter label.

Stations in the inside waters of Southeast Alaska (Area 2C) were not laid out according to a geometrical pattern but were distributed purposively over several of the major inside grounds. These stations were assigned individual station numbers in the range 401 through 824, and no letter label.

Table A1 lists all of the stations that constitute the IPHC grid survey data, plus the inside stations in Area 2C, for the years 1963-1986. Note that stations with missing and zero station numbers have to be excluded when the data are retrieved.

Other selection criteria

Station type

Users interested in truly systematic stations only may want to exclude the stations in the inside waters of Area 2C in the early data (1963-1986), because they were placed arbitrarily. That can be done on the basis of station number. In the recent data (the 2C series resumes with 1996), the inside stations with a purpose code of "SG" are in fact regular grid stations. The survey vessels sometimes fished additional stations for operational reasons, but those stations have a purpose code of "ES".

In 1993, offshore stations were placed in groups of four, of which three were the vertices of triangles laid out systematically on a grid. The skipper of the survey vessel was free to choose the location of the fourth station anywhere within the triangle, so those stations are not truly

Table A1. List of all IPHC grid survey stations through 1986. Some stations fished by the listed vessels have no station numbers at all; these are not survey stations and must be excluded.

Surveys in Canada

Year	Vessels	Stations
1965	CHR	Stations with numbers > 0
1966	CHE	Stations with numbers > 0
1976	SEY	Stations with numbers > 0
1977	CHE, EVE	Stations with numbers > 0
1978	CHE	Stations with numbers > 0
1980	ELL	Stations with numbers > 0
1981	PRC	Stations with numbers > 0
1982	PRC	Stations with numbers > 0
1983	EVE, WIN	Stations with numbers > 0
1984	STW, WIN	Stations with numbers > 0
1985	STW	Stations with numbers > 0 and < 900
	CFL	Stations with numbers > 0
1986	SNO, WIN	Stations with numbers > 0

Surveys in Alaska

Year	Vessels	Stations
1963	ECL	Stations with numbers > 0
1964	ECL	Stations with numbers > 0
1965	CHE	Stations with numbers > 0
1976	POL	Stations with numbers > 0
1977	POL, RES	Stations with numbers > 0
1978	VAN	Stations with numbers > 0
1979	CHE	Stations with numbers > 0
1980	SEY	Stations with numbers > 0
1981	EVE	Stations with numbers > 0
1982	KRI, THR	Stations with numbers > 0 (fixed gear)
	DLY, VAL	Stations with numbers > 0 (snap gear)
1983	POL, MAS	Stations with numbers > 0 (fixed gear)
	VAL	Stations with numbers > 0 (snap gear)
1984	CHE, SEY	Stations with numbers > 0
1985	CHE	Stations with numbers > 0 and < 900
	CFL	Stations with numbers > 0
1986	CFL	Stations with numbers > 0

systematic, either. They are distinguished in the 1993 data by a two-letter rather than a one-letter station label ("station position" in the database), formed by appending "C" to one of the regular one-letter labels. As a practical matter the catches at the skipper stations were substantially higher than at the other stations in 2B (9.8 vs. 5.5 fish/skate, P=0.001) but slightly lower in 3A. In 1994 and 1995 the same station pattern and coding scheme were used, but the interior station was always placed at the centroid of the triangle.

Gear type

The standard gear for surveys has always been fixed (stuck) gear (gear code "FH"), but in three cases—noted in Table 1—vessels fished snap gear (code "SN") for comparative purposes. In all cases the vessel using snap gear fished a subset of the stations fished by another vessel using fixed gear in the same area and year, so the snap gear sets can all be excluded without loss of coverage, and should be excluded to avoid giving extra weight to the subset of stations fished twice. At the 139 stations fished with both gears in 1982 and 1983, fixed gear catches were significantly higher than snap gear catches, but not by a large amount (10.3 vs. 8.7 fish/skate, P=0.0001).

Hook type

Surveys were done with J-hooks (hook type code "J") through 1984. In 1984 and thereafter C-hooks (code "C") were used. In 1984 all of the stations in Areas 2B and 3A were fished with both J- and C-hooks for comparative purposes. Five survey stations in 1984 and 1985 were fished with a mixture of J- and C-hooks; they have a code of "M" and should normally be excluded.

Effectiveness of set

Beginning in 1993, each set was assigned an "effectiveness" code, either "Y" for a normal, successful set or "N" for an unsuccessful one (e.g., lost most of the gear, set in the wrong place etc.) . Most users of the data will want to exclude unsuccessful sets.

For years before 1993 the effectiveness code is empty, but there are a few sets where the number of *effective* skates hauled ("effskt") is given as zero even though it is clear that some gear was hauled (e.g., in very rough weather when many fish are lost during hauling). These sets should also be regarded as ineffective.

Number of skates hauled

The number of skates set at each station has varied over the years, the minimum being the equivalent of four 100-hook skates. Part of a string is often lost, so the number of skates hauled can be less. It makes sense to set some minimum number of skates hauled in a data retrieval to avoid catch rates based on a small number of hooks. Setting the minimum at two 100-hook skates is reasonable and eliminates only a small proportion of sets.

In the database, the field called "number of skates" ("noskt") is the number of skates actually hauled, but this does not refer to standard 100-hook skates with 18-foot spacing. Skate length and hook spacing both vary among trips. Surveys in the 1960s and 1990s were all done with 18-foot gear; almost all surveys in the 1970s and 1980s with 21-25-foot gear. Other things being equal, catch per hook is higher if hook spacing is larger. The relationship has been studied (Skud 1972), so the effect of non-standard hook spacing can be removed by calculating an adjustment.

A field in the database called "effective skates" ("effskt") is the number of skates hauled, adjusted for hook type, hook spacing, and skate length to the equivalent number of standard skates of 100 C-hooks at 18-foot spacing. The number of C-hook skates is not adjusted for hook type, but the number of J-hook skates is divided by 2.2 because C-hook catches of legal-sized

Table A2. Extent of survey coverage by year. The core survey area in 2B is the part north of Vancouver Island (Fig. 2); the core area in 3A is the western part of the outer shelf (Fig. 3). A grid survey was first conducted in 2A in 1999.

Year	Area 2B	Area 2C	Area 3A	Area 3B	Area 4
1963	---	---	Core	---	---
1964	---	---	Shelikof Strait only	Eastern half only	---
1965	Southern part of core	---	---	Western half only	Eastern 4A only
1966	Northern part of core	4 stray stations	---	---	---
1976	Northern part of core	---	Only 2 transects	---	---
1977	Core	---	Core	---	---
1978	Core	---	Core	---	---
1979	---	---	Core	---	---
1980	Core (fewer stations)	---	Core (fewer stations)	---	---
1981	Core	---	Core	---	---
1982	Core	All[1]	Core	Western half only	---
1983	Core	All[1]	Core	---	---
1984	Core (twice, C&J)	All[1]	Core (twice, C&J)	---	---
1985	Core	All[1]	Core	---	---
1986	Core	All[1]	Core	---	---
1993	Core	---	Core	---	---
1994	---	---	Core	Eastern half only	---
1995	Core	---	Core	3 stray stations	---
1996	Core	All	All outside waters	All	---
1997	Core	All	All outside waters	All	4A, 4B, 4C, 4D edge
1998	Core	All	Outside + inside waters	All	4A, 4B
1999	All	All	Outside + inside waters	All	4A, 4B

[1] Stations in inside waters chosen purposively.

fish in weight were 2.2 times J-hook catches in paired sets in 1984. The relative fishing power of the two hook types actually depends strongly on the size of the fish (Sullivan et al. 1999), so the adjustment is not really reliable.

The number of standard 100-hook skates actually hauled, adjusted for hook spacing and skate length but not hook type, can be obtained by taking the "effskt" value for C-hooks, and multiplying that value by 2.2 for J-hooks.

As explained above, sets where "effskt" is zero should be regarded as ineffective regardless of the value of "noskt".

Survey area

Before 1999, grid surveys in Area 2B were confined to the waters north of Vancouver Island (Fig. 2), and in some years did not cover all of those waters. In 1999 all of 2B was surveyed, and for the first time 2A as well. (Random stratified surveys, with high sampling densities on commercial grounds, were carried out in 2A and southern 2B in 1995 and 1997.) Likewise in Area 3A, only the western part of the shelf (west of 148ºW, excluding Cook Inlet and Shelikof Strait; Fig. 3) was normally surveyed before 1996, and in some years not all of that. In 1996 the eastern half of 3A was added, and in 1998 stations in Prince William Sound, Shelikof Strait, and lower Cook Inlet were added. Area 2C was always surveyed entirely if at all, but only in recent years (1996-) have regular grid stations been fished in the inside as well as the outside waters. For Area 3B and Area 4 there are only spotty data before comprehensive surveys began in 1996 (3B) and 1997 (4). The evolution of survey coverage over time is summarized in Table A2.

Assembling a consistent series of survey data for 2B requires filtering out the years with only partial coverage of the core survey area. It also requires setting some northern boundary for Area 2B, which is controversial because Canada and the U.S. both claim a large part of the northern side of Dixon Entrance. The operational boundary used in designing the surveys has usually been 54º30'N. Beginning with 1999 it is also necessary to filter out stations off Vancouver Island (south of, say, 50º30'N). Assembling a consistent series of grid-only data for Area 2C requires filtering out all of the inside stations. That can be done on the basis of station number for years through 1986, but thereafter only on the basis of latitude and longitude. Assembling a consistent series for 3A requires filtering out years with partial coverage of the core area before 1996.

References

Hoag, S. H., Williams, G. H., Myhre, R. J., and McGregor, I. R. 1980. Halibut assessment data: setline surveys in the North Pacific Ocean, 1963-1966 and 1976-1979. Int. Pac. Halibut Comm. Tech. Rep. 18.

Randolph, D. L. 1998. IPHC setline survey grid design changes 1960-1998. Int. Pac. Halibut Comm. Report of Assessment and Research Activities. 1998:293-300.

Skud, B. E. 1972. A reassessment of effort in the halibut fishery. Int. Pac. Halibut Comm. Sci. Rep 54.

Sullivan, P. J., Parma, A. M., and Clark, W. G. 1999. The Pacific halibut stock assessment of 1997. Int. Pac. Halibut Comm. Sci. Rep. 79.

Checklist

Here is a checklist of selection criteria that should be considered when retrieving grid survey data:

Feature	Criteria
Station number	Valid number: not missing or zero
Vessel, year, station number	Table 1 for years 1963-1986. Purpose code = “SG” for years 1993 on.
Station type	Exclude 2C inside stations before 1993? Exclude skipper stations in 1993?
Gear type	Exclude snap gear (gear code = “SN”)?
Hook type	Exclude sets with mixed hooks (hook code = “M”)? Select only J-hook (code “J”) or C-hook (“C”) data?
Effectiveness of set	Exclude sets with zero effective skates. Exclude ineffective sets (code = “N”)? (Effectiveness field is empty for years before 1993.)
Number of skates hauled	Exclude sets with too few skates hauled?
Area coverage	Exclude years with partial coverage (Table 2)? Exclude stations outside core areas in 2B and 3A?

Appendix B. A method of estimating the sex composition of commercial landings from setline survey data

This is a condensed version of:
Clark, W. G. 2004. A method of estimating the sex composition of commercial landings from setline survey data. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2003: 111-162.

The whole text of the paper is reproduced here. The original paper has many more figures showing the data and fits to support statements made in the text.

Abstract

The sex ratio of fish of a given age and length in survey catches does not vary appreciably with depth or with location within a regulatory area in any given year, so it seems safe to assume that commercial landings of fish in the same category have the same sex ratio as survey catches. The relationship between length and sex ratio (proportion female) for fish of a given age does vary among regulatory areas and years, but in all cases it is well described by a simple logistic function. For recent years, the commercial age/length samples can therefore be keyed out to sex by estimating the two logistic parameters for each area/year/age category from the survey data. For earlier years when the actual lengths of fish in the commercial landings were not measured, the age-specific sex and size compositions of legal-sized fish in the survey catches appear to be reasonable estimates of the commercial values.

Introduction

Setline gear is selective for larger fish, and the commercial fishery has an 81 cm minimum size limit. Female halibut are larger at each age than males, so they presumably sustain a higher fishing mortality. The dramatic reduction in halibut growth rates over the last fifteen years, especially in Alaska, has likely increased the difference in fishing mortality between the sexes.

Because halibut are eviscerated at sea, we have no sample data on the sex composition of commercial landings, so the annual stock assessment has never distinguished the sexes. Instead a single catchability coefficient and selectivity schedule have been calculated for the combined sexes, and both exploitable and spawning biomass have been computed from estimates of combined abundance.

Concern about the actual fishing mortality rate of females and the actual level of female spawning biomass relative to earlier levels led to a decision to attempt an age- and sex-structured assessment in 2003; i.e. to use a model that does the usual population calculations for females and males separately. Sex-specific catchability and selectivity parameters can be estimated for the survey because we have a sexed sample, but the commercial data are a problem.

One approach is to continue estimating a single commercial catchability coefficient and age-specific selectivity schedule. Because those parameters are modeled as random walks, they could in principle track changes in the age-specific averages due to changes in age-specific sex composition. But that would mean applying the same fishing mortality rate to females and males at each age in the population accounting, which is almost certainly wrong.

Another possibility is to estimate a single commercial catchability coefficient and length-specific selectivity schedule (both random walks) and then use sex-specific mean length at age in the survey to predict fishing mortality and therefore catches at age by sex. This approach requires the strong assumption that the same length-specific values apply to both sexes, and even if that is true, it is doubtful that the parameters would be estimable in the absence of observations of catch at age by sex. The resulting estimated commercial sex compositions could also turn out to be quite different from the observed survey sex compositions, which would be problematic.

The simplest solution is to somehow estimate commercial sex compositions from the survey sex compositions external to the assessment and then do straightforward, mostly parallel age-structured assessments of each sex. Sex-specific commercial parameters would then be readily estimable, and the working estimates of commercial sex composition would agree with the survey data. This paper reports a method of estimating the sex composition of the commercial catch at age using a survey-based estimate of the age-specific relationship between length and sex ratio.

Mean length at age in survey catches and commercial landings

Because females are larger than males at each age, the mean length at age in commercial landings is an indicator of sex composition. In particular, if the mean length at age in commercial landings closely matched the mean length at age of legal-sized fish in survey catches, that would be strong evidence that commercial and survey sex ratios at age (and sex-specific length distributions at age) are the same. This appears to be the case in Area 2B (Fig. B1a), and until the last few years in Areas 2C and 3A. But in Areas 3B (Fig. B1b) and 4 the mean length at age has always been much larger in commercial landings than in survey catches, and the same pattern has appeared in Areas 2C and 3A recently.

There are a number of possible reasons why the commercial fishery could select larger fish at each age than the survey, including ground selection, gear differences, and highgrading. Whatever the reason, selecting larger fish must result in a higher proportion of females in commercial landings, at least in some areas and years.

Length frequencies at age in survey catches and commercial landings

For all survey years back to 1993, when surveys resumed after a 6-year hiatus, it is possible to compare age-specific commercial and survey length frequencies to see what causes the differences in mean age at length. In the survey data the female and male components of each length interval can also be distinguished so as to show what sort of differences between survey and commercial sex composition might result from the differences in length composition.

Direct comparisons of length frequencies cannot be made for earlier years because the fork lengths of fish in commercial samples were not measured between 1963 and 1990. Instead one or another measurement of the otoliths was made and fork lengths were calculated from a predictive relationship. At best this sort of prediction would provide the expected fork length of each fish, but the distribution of expected fork lengths would not be the same as the distribution of true fork lengths. To further complicate matters, the relationship between otolith size and fork length changed over time. The staff has reworked the predictions to improve the estimates of mean size at age in commercial landings (Clark 1992), but there is still some doubt even about the means, and there is no satisfactory way to recover the entire length frequency distributions.

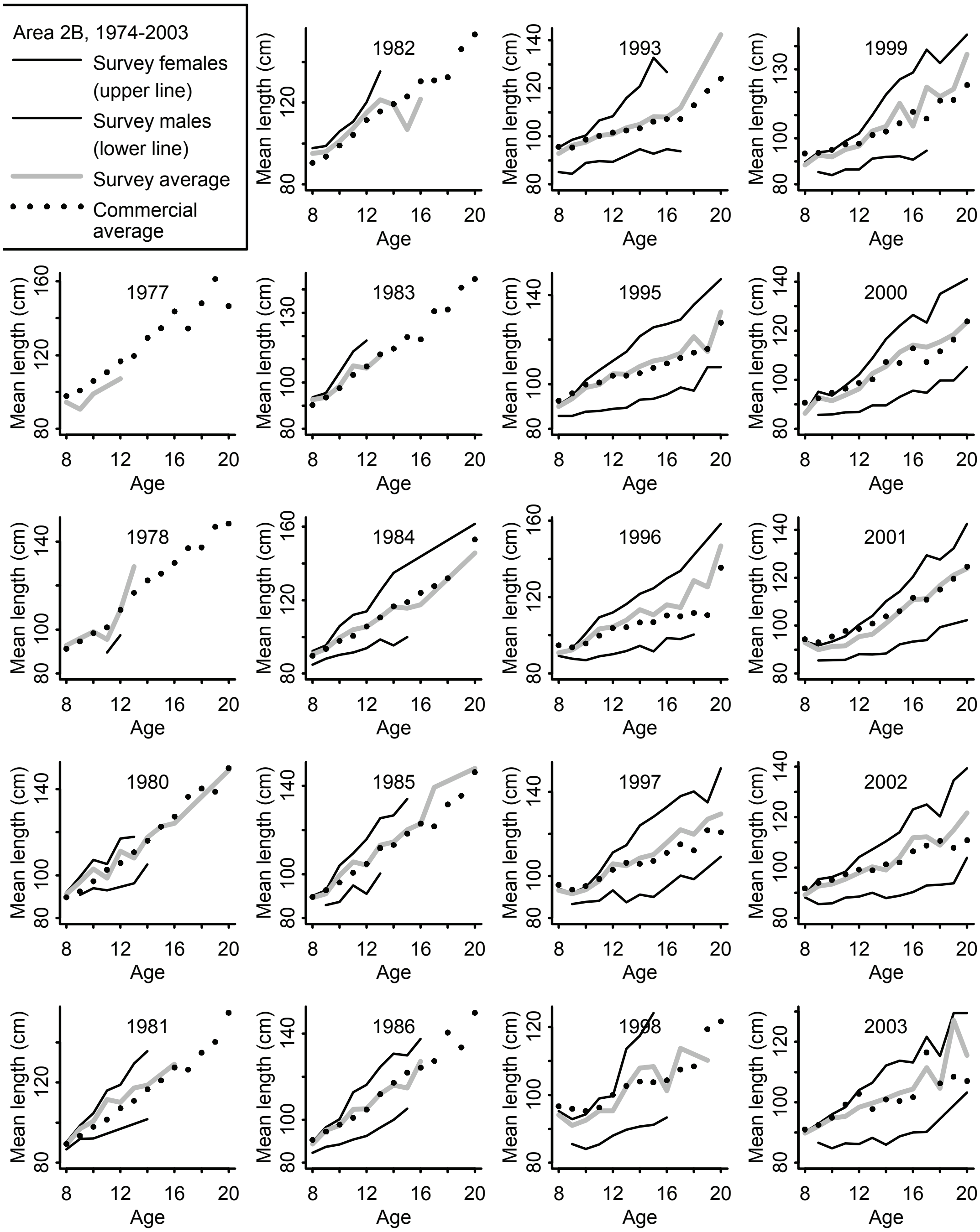


Figure B1a. Mean length at age of legal-sized fish in commercial landings and survey catches in Area 2B.

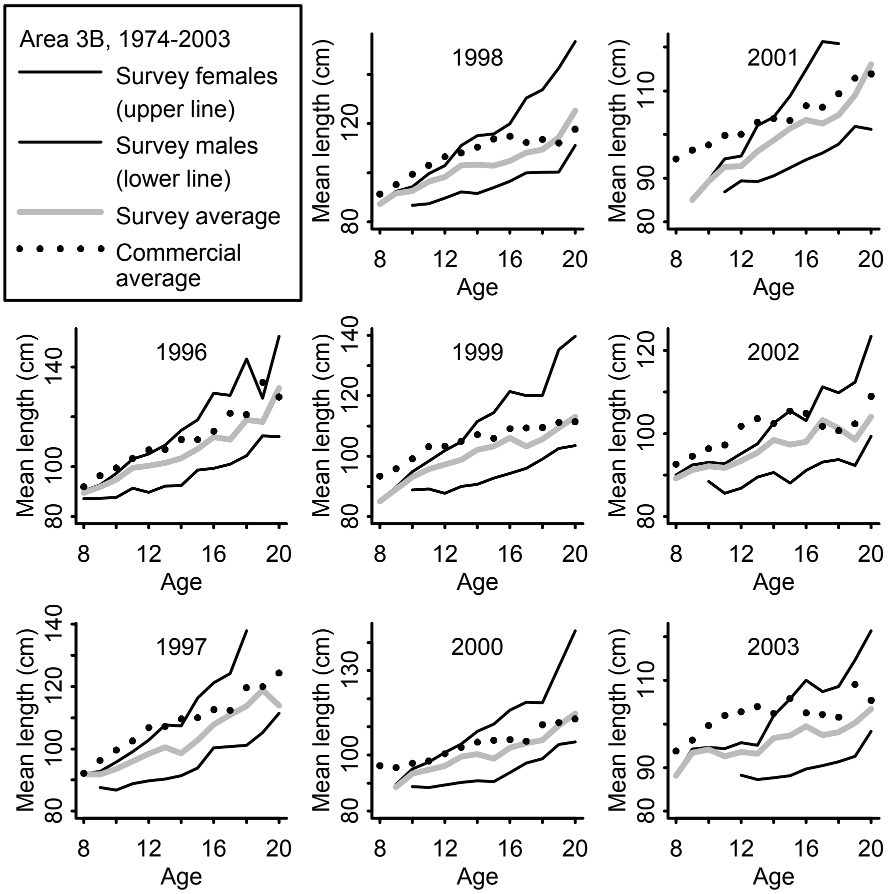


Figure B1b. Mean length at age of legal-sized fish in commercial landings and survey catches in Area 3B.

The available data are further complicated by a change in age reading practices at IPHC. All of the age data in the database through 2001 are surface readings, which on average underestimate the ages of fish older than 12 or so. Beginning with 2002 all of the data are break-and-burn readings, which are accurate. For that reason the analysis is this paper will generally treat the age data types separately, meaning that most comparisons show data from the more numerous years of surface age readings, ending in 2001. The methods eventually developed apply equally well to 2002 and later data, as is shown, and are applied to those data separately to obtain working formulas.

In recent years where direct comparisons between survey and commercial data can be made (i.e., surface reading years when surveys were done and commercial lengths were measured), the data show that relative to the survey length frequencies, the commercial landings contain substantially fewer fish under 90 cm and slightly more fish across a wide range of greater lengths, where apart from that slight difference the distributions are usually similar. This is true even in those cases where survey and commercial mean lengths at age agree closely, namely in Area 2B (Fig. B2a) and until 2000 in Areas 2C and 3A. The pattern is simply more pronounced in Areas 3B and 4 where the commercial mean lengths are consistently higher (Fig. B2b). Interestingly, limiting the survey length frequencies to data from stations deeper than 100 fm largely eliminates the pattern in Area 3A and to a lesser extent in other areas, suggesting that it results at least in part simply from the difference in grounds fished by the survey and the commercial fishery.

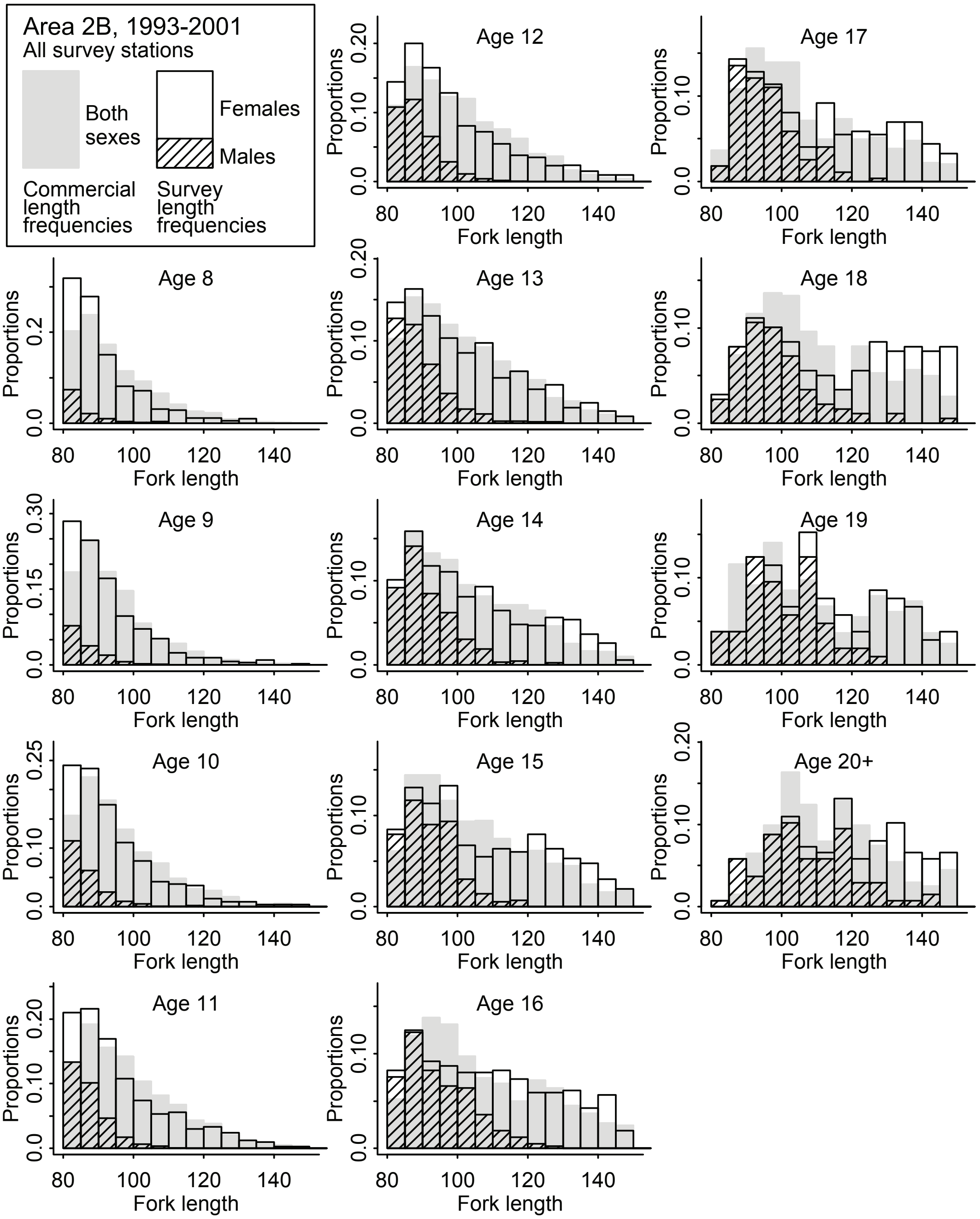


Figure B2a. Survey and commercial length frequencies at age in Area 2B, 1993-2001 data pooled. Survey and commercial mean length at age agreed well in all these years.

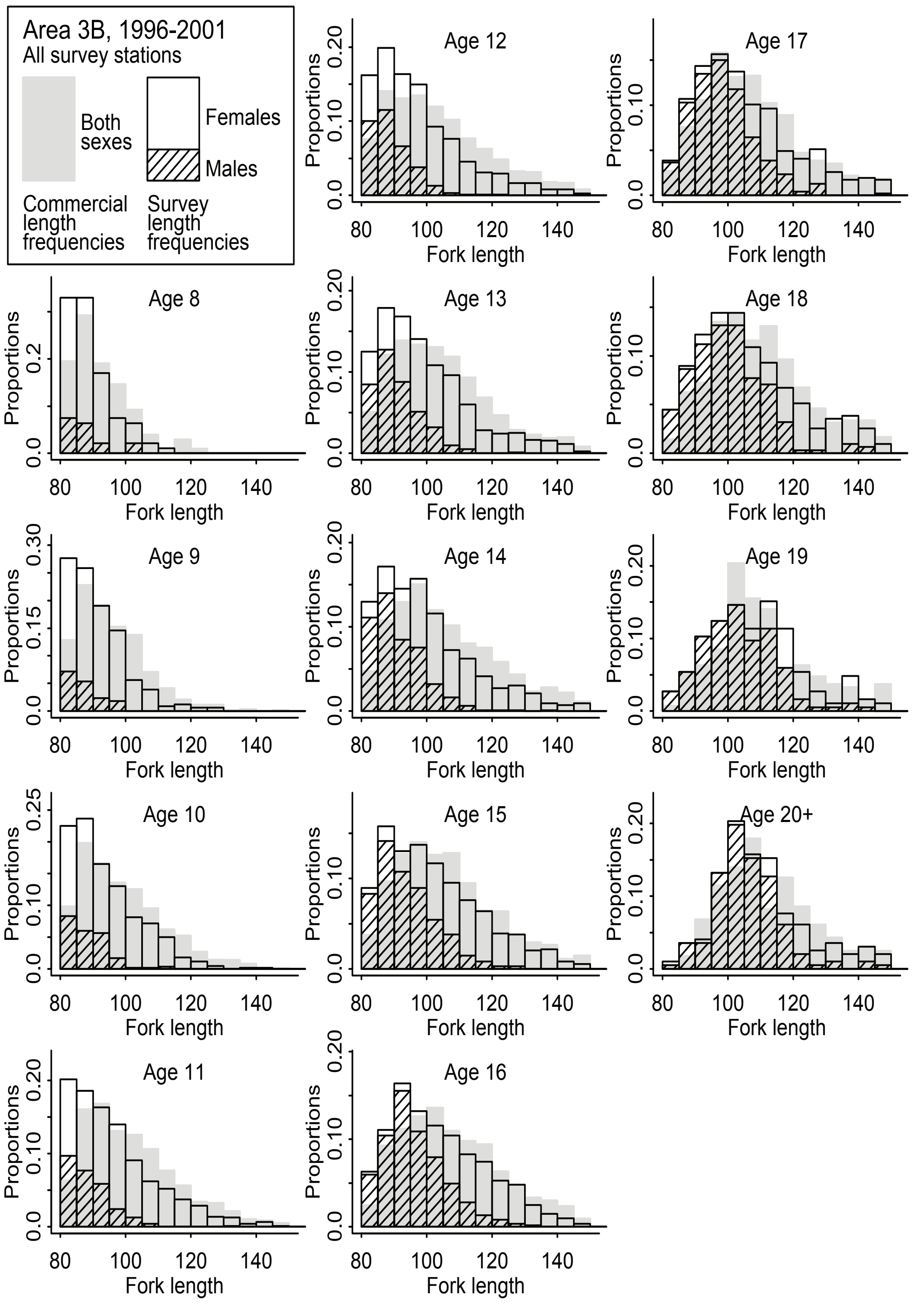


Figure B2b. Survey and commercial length frequencies at age in Area 3B, 1996-2001 data pooled. Commercial mean lengths at age exceeded survey means in these years.

At every age the 80-90 cm length group contains a substantially higher proportion of males than the age group as a whole. The commercial landings have relatively fewer fish of this size than the survey catches, so they presumably contain a higher proportion of females than the survey catches. The commercial sex ratio can be estimated by applying the survey sex ratio to the commercial frequencies in each length interval, but only if the sex ratio in an age/length category does not vary among grounds within a regulatory area.

It is shown below that this is the case, which is somewhat surprising given the difference in size between females and males, and the difference in length compositions between shallow and deep survey stations. But the data do in fact show little difference in sex ratio at a given age and size between deep stations and all stations in any area, and plots of the depth distributions of females and males of the same age are quite similar at all ages in all areas with the exception of teenaged fish in Area 2B, where the males tend to be shallower (Figs. B3a and B3b). Younger fish do have a more inshore distribution in all areas, but it appears that both males and females move deeper with age. Another confirmatory piece of evidence is that the sex ratio at a given length is the same in trawl and setline survey catches in Area 3A (Clark 2001) despite large differences in overall length composition and mean length at age.

Direct estimates of the sex composition of commercial catches

This section of the paper reports estimates of the sex composition of commercial landings computed by applying the raw survey values of proportion female in each age/length category to the age/length frequencies of the commercial data, with enough years of data pooled to avoid small-sample problems. This is the natural way of computing the estimates, and it will show in a simple and direct way the size of the differences between survey and commercial sex composition in the various regulatory areas in recent years. The following sections of the paper will use fitted curves to represent the relationship between length and sex ratio within an age group so as to simplify comparisons and eventually working calculations.

Applying the survey sex ratio (proportion female) in each length interval to the commercial length frequencies for a given age group provides estimates of the commercial sex ratio at that age and the size distributions of females and males of that age in the landings, all quantities of interest for purposes of stock assessment and harvest policy evaluation. These estimates show that for Area 2B (Fig. B4a), and until recently Areas 2C and 3A, the overall sex compositions of commercial landings were very similar to the sex compositions of legal-sized fish in survey catches, despite the relative paucity of 80-90 cm fish in the landings. In these cases the estimated proportion female and mean size at age by sex are very slightly higher in the commercial estimates, but all of the differences are small compared with the year-to-year variability of survey values and, for years before 1991, the uncertainty concerning the true mean size at age in commercial landings. In view of this finding, and the generally good agreement between survey and commercial mean size at age going back to 1974, it appears to be practical to use the observed sex composition of legal-sized fish in the survey catch as estimates of the commercial values in Areas 2B, 2C, and 3A from the 1970s through the late 1990s.

The commercial estimates of proportion female at age in recent years are consistently and substantially higher than the survey values in Areas 3B (Fig. B4b) and 4A, and for most ages in Area 4B. It will therefore be necessary to key out the commercial length frequencies to estimate the commercial sex composition, effectively limiting the stock assessment in those areas to the beginning of systematic surveys in 1996-97.

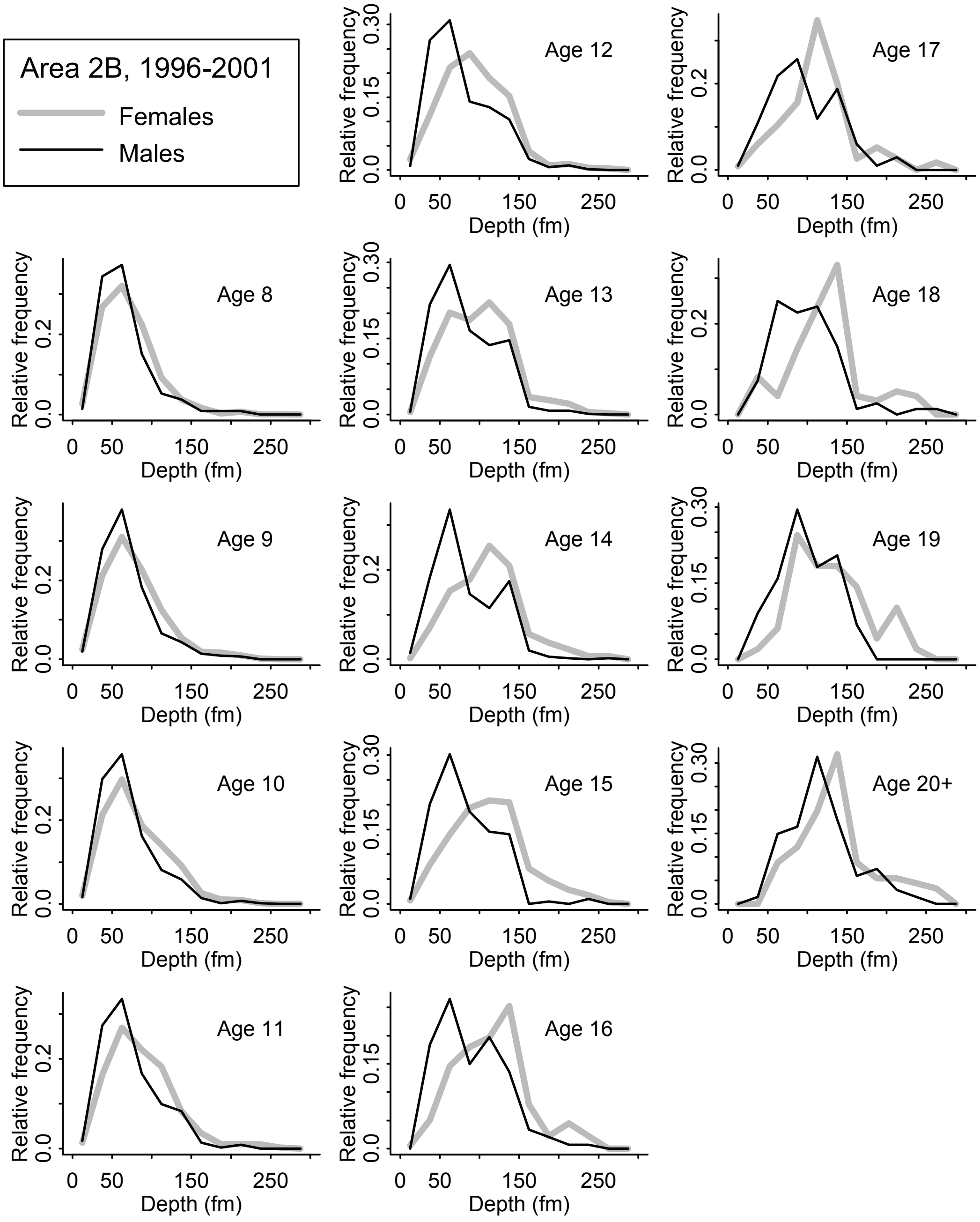


Figure B3a. Depth distributions of survey catches of females and males of the same age in Area 2B.

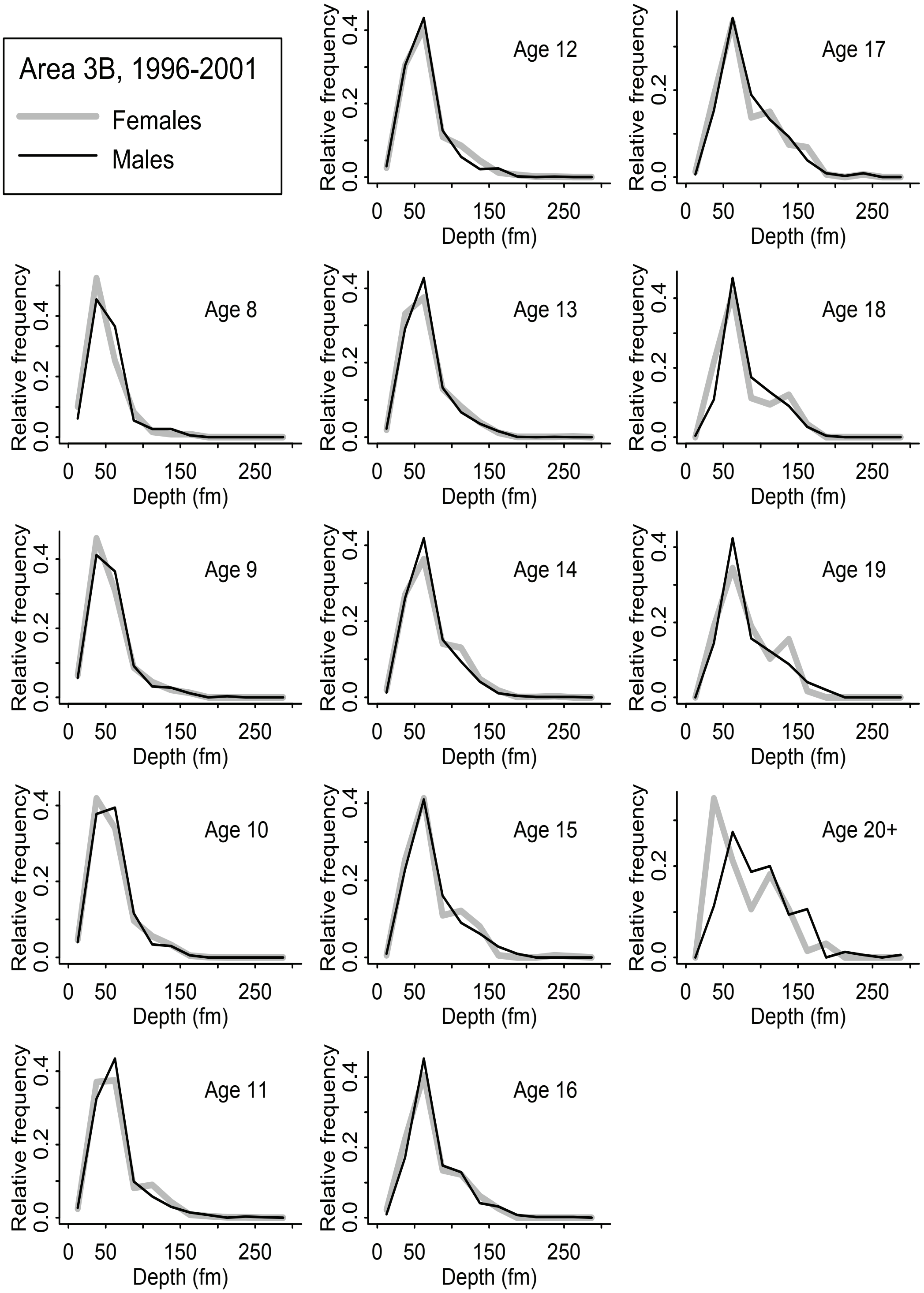


Figure B3b. Depth distributions of survey catches of females and males of the same age in Area 3B.

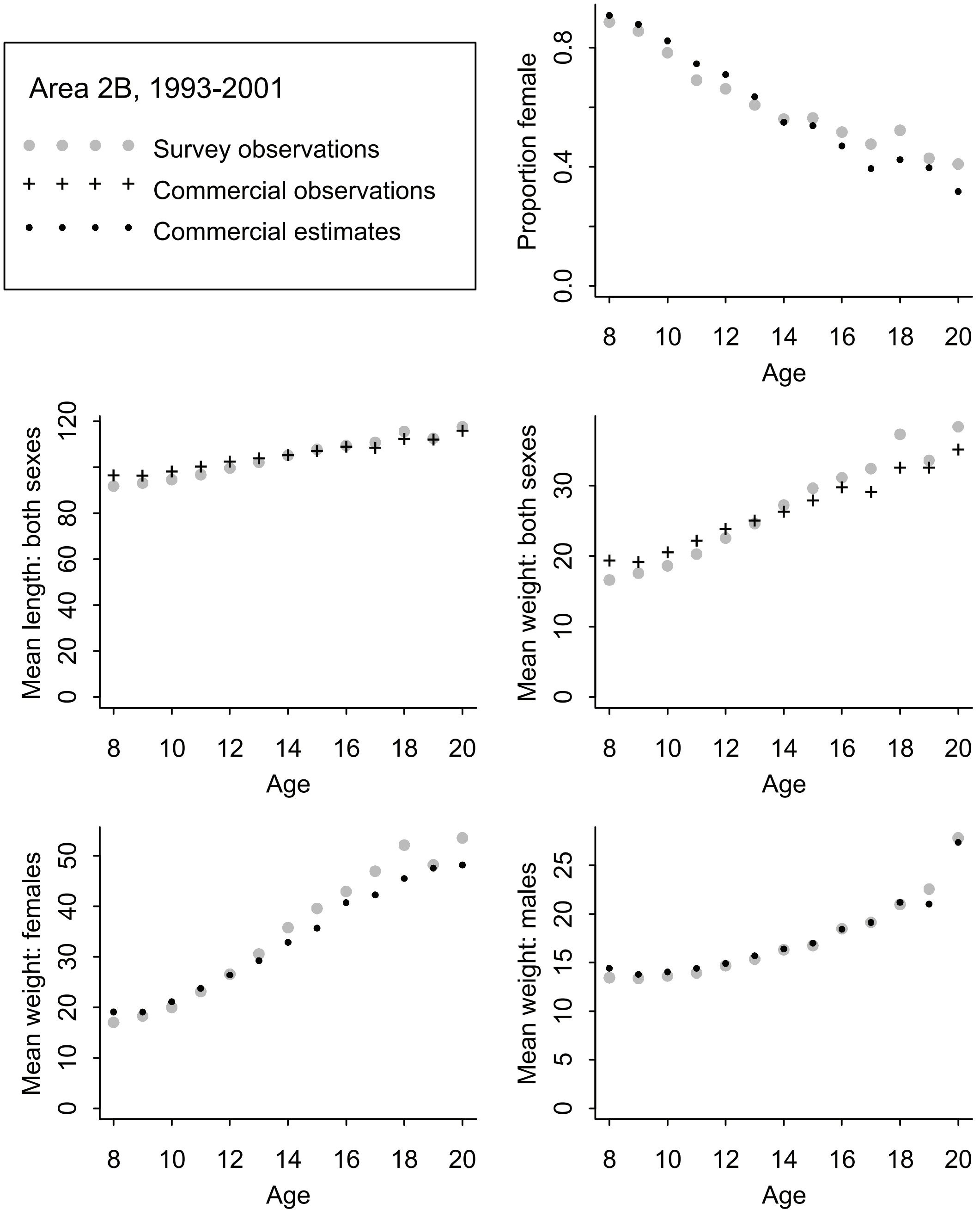


Figure B4a. Estimates of proportion female and mean size by sex in commercial landings in Area 2B, 1993-2001 combined, computed for each age by applying the sex ratio at length in survey catches to the length frequencies of commercial landings.

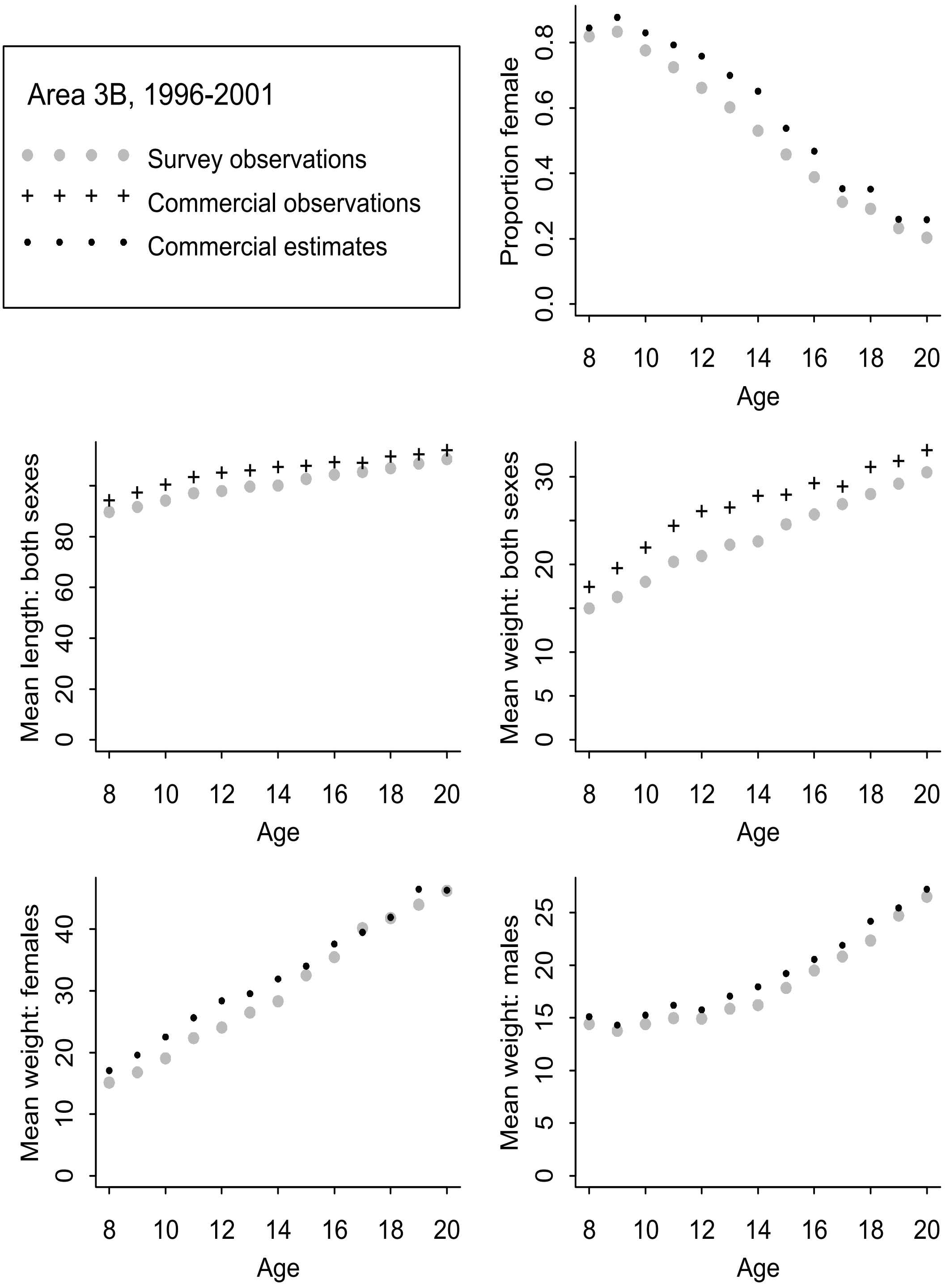


Figure B4b. Estimates of proportion female and mean size by sex in commercial landings in Area 3B, 1996-2001 combined, computed for each age by applying the sex ratio at length in survey catches to the length frequencies of commercial landings.

Form of the relationship between sex ratio and length at a given age

It remains to be shown that the sex ratio at a given age and length is the same throughout a regulatory area in any given year. Comparisons between subareas and depth zones can be simplified by fitting a curve for each case of interest and then comparing the fitted curves. The relationship between proportion female p and fork length L for fish of a given age in a given area in a given year (all fish, not just legal-sized ones) is very well described in all cases by a simple logistic function:

$$p = 1 \Big/ \left(1 + \exp\left(-k \cdot \left(L - L_{50}\right)\right)\right)$$

where k is a slope parameter and L_{50} is the length at which 50% of the fish are female (Fig. B5). The location and steepness of this curve must depend in a very complicated way on the growth and exploitation histories of the females and males in the age group, so there is no question that the parameter values will vary among areas and years. But the form is remarkably consistent.

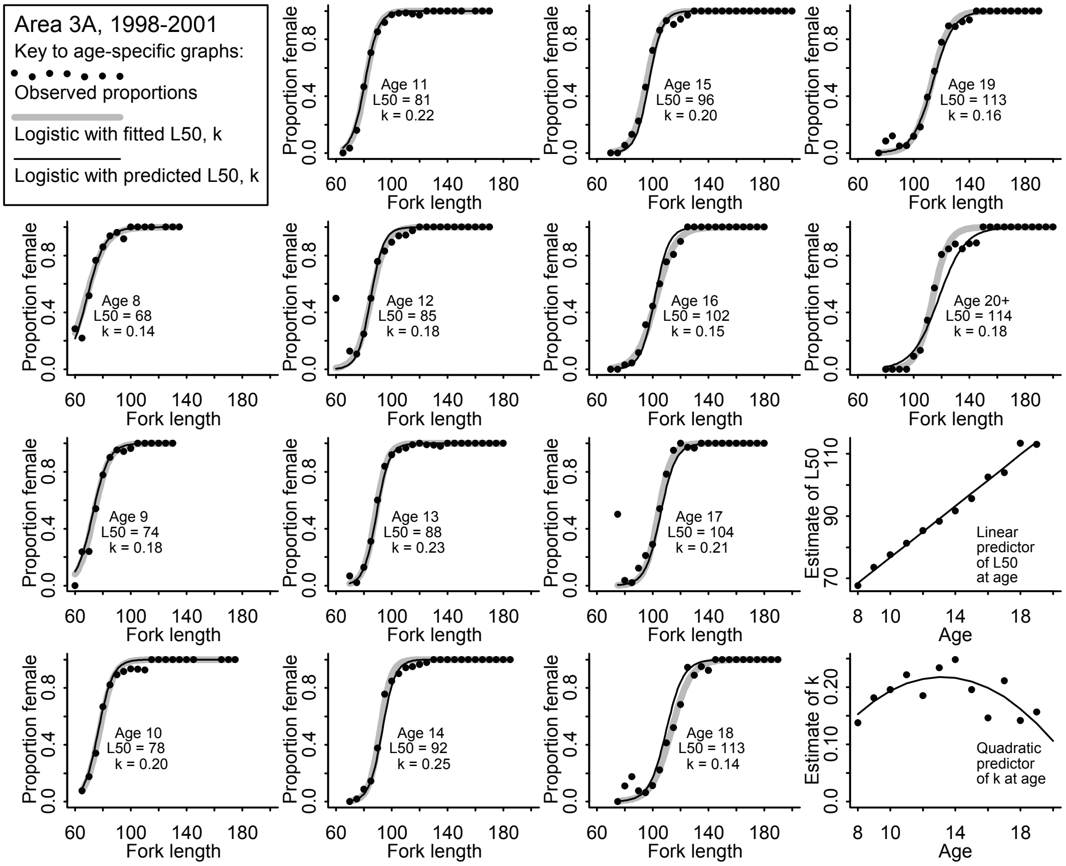


Figure B5. Logistic fits to the proportion female at length, by age, in C-hook survey catches in Area 3A, 1998-2001.

Another consistent feature of the relationship is a very tight linear trend with age in the estimate of L_{50} for the fish in a given area and year (second to last panel in Fig. B5). The estimates of *k* are more variable from one age to the next and sometimes exhibit a modest curvature (last panel in Fig. 5). The variation of L_{50} with age can be represented very well by a fitted straight line:

$$L_{50} = b_0 + b_1 \cdot age$$

and the variation of *k* with age can be represented well enough (the fits are not very sensitive to the precise value of *k*) by a fitted quadratic:

$$k = c_0 + c_1 \cdot age + c_2 \cdot age^2$$

Logistic functions with parameters predicted in this way closely match the age-specific fitted logistics for all ages except the plus group (e.g., 20+ for surface ages), which consists of a mixture of older age groups and therefore tends to have a higher L_{50} and lower *k* than predicted by the trends among the younger age groups. To specify the proportion female as a function of length for all ages in a given area and year therefore requires seven meta-parameters: the coefficients b_0, b_1, c_0, c_1, c_2 shown above and estimated from the trends in L_{50} and *k* among all ages except the plus group, and the values of L_{50} and *k* for the plus group.

The comparisons in the next sections show separate logistics fitted to the data for each age group. Meta-parameter values are computed later in the paper for use in working calculations.

Consistency of the relationship between depth zones and subareas

Because survey and commercial effort are distributed differently, a single survey-based schedule of sex ratio at length for a given area/year/age will be correct only if the schedule is the same throughout the area, i.e. in all depths and all subareas.

When the survey data are divided in two by depth, the separate logistics fitted to the shallow and deep components are very close in all cases. The largest differences occur in Area 2B, which as noted above appears to be the only area where there is any difference in the depth distribution of females and males of the same age. But even in Area 2B, the relationship between length and sex ratio at a given age is almost the same for shallow and deep stations, at least through age 16. Beyond that there are not enough old fish at shallow stations to make the comparison using only 2B data, but using data from all of Area 2 shows reasonably close agreement among the older fish as well. This is true both for the years with surface age data (through 2001), and for 2002-2003 when all the age data are break-and-burn readings. In other regulatory areas the agreement between shallow and deep fits is as good as or better than in Area 2.

In similar comparisons, there is no difference in the fits between the southern and northern halves of 2B or 2C, or between the eastern and western halves of 3A, 3B, 4A, or 4B. This is not to say that there are no differences at all within regulatory areas, but only that they are small enough to be disregarded.

There are some large differences between regulatory areas. In recent years there has been little difference between Areas 2B and 2C, or between Areas 2A and 2B, but some noticeable and consistent differences between Areas 2C and 3A; also between Areas 3A and 3B. Areas 3B and 4A are almost identical, but there are again large differences between Areas 4A and 4B. So while a single fit is sufficient for an entire regulatory area, it is desirable to treat regulatory areas separately, and there is no reason not to.

Changes in the relationship over time

Because of the large changes in growth schedules that have occurred in the last twenty years, the schedule of sex ratio at length has shifted toward smaller sizes, modestly in Area 2B where the growth changes were less (Fig. B6a) and dramatically in Area 3A (Fig. B6b). In the mid-1980s, the length at which the proportion female reached 50% among the modal age groups (ages 10-14) in Area 3A was 120-140 cm; now it is 80-90 cm. The bulk of this shift occurred between the mid-1980s and the mid-1990s; during the latter 1990s there was some small further movement, particularly in the 20+ age group, but not a great deal.

Working formulas

In 1996 setline surveys were begun in Area 3B, extended eastward in Area 3A to cover nearly the whole area, and resumed in Area 2C. It is therefore a convenient starting year for estimating the sex composition of the commercial landings using the procedure developed above. Surveys in Areas 4A and 4B did not begin until 1997, but because the sex ratio at length (within age) does not appear to have been changing rapidly in the latter 1990s, it should be possible to key out the 1996 commercial data in those areas reasonably well using the 1997-1998 survey data. Commercial landings in Area 2A can be keyed out with Area 2B survey data at need.

For each area and each year beginning with 1996, the seven meta-parameters of age-specific logistic functions described above have been computed. This was done using a moving three-year data window to calculate the estimates for the central year, except at the beginning and end where only two years of data were used. Where fewer than 100 observations were available at a given age, data were added from neighboring ages (to a maximum of ±2 years) to reach a minimum sample size of 100 for each age-specific estimate. Extreme outliers from a fitted logistic were removed and the curve refitted to improve the parameter estimates. The data series were broken between 2001 and 2002 at the time of the change from surface to break-and burn readings because that did affect the shape of the fitted logistics. The plus age is 20 for the surface years and 25 for the break-and-burn years. Only in Area 4 did fish over age 25 represent more than 1% of the landings—2% in Area 4A, 7% in 4B, and 5% in 4D.

For years before 1996, smoothed survey values (for legal-sized fish) were used to estimate the age- and sex-specific commercial values in Areas 2B, 2C, and 3A. This involved a long interpolation between 1986 and 1993 when surveys were suspended, but in fact this was a period of more or less steady decline in estimated commercial mean length at age, so the linear interpolation should be reasonably accurate.

Results

Application of the estimation procedure described above to commercial length frequencies produces estimates of sex and size frequencies that are very consistent with the observed characteristics of legal-sized fish in the survey in the mid-1990s. Mean size at age by sex and sex ratio at age show good continuity from the period where raw smoothed survey observations are used (1974-1995) to the period where the sex and size compositions are estimated from the commercial length frequencies at age (1996-2003). This is not surprising, since the logistic predictors were estimated from the survey data. What the good agreement shows is that the use of meta-parameters to summarize the logistics does not degrade the estimates.

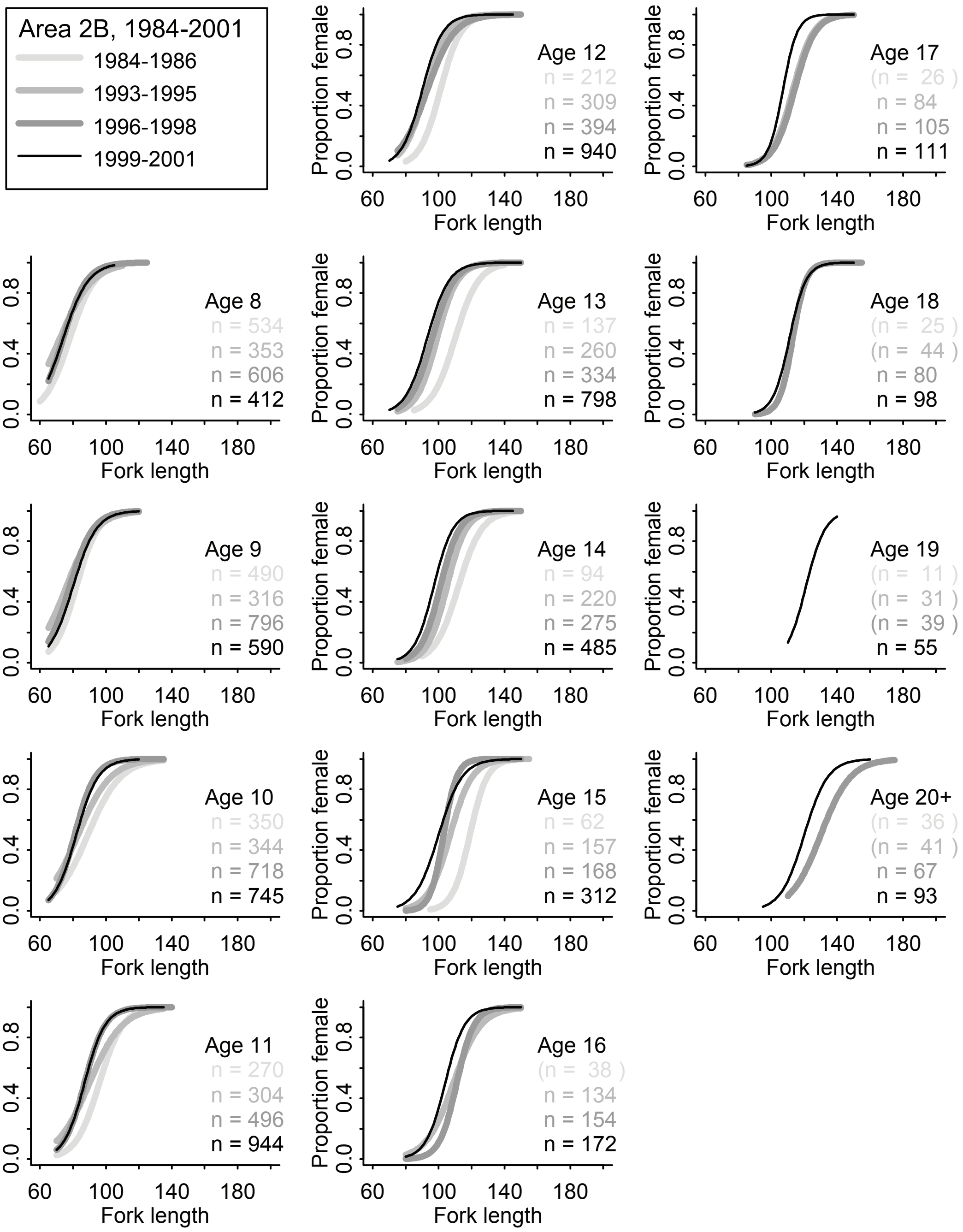


Figure B6a. Logistics fitted to proportion female at length in Area 2B by three-year periods.

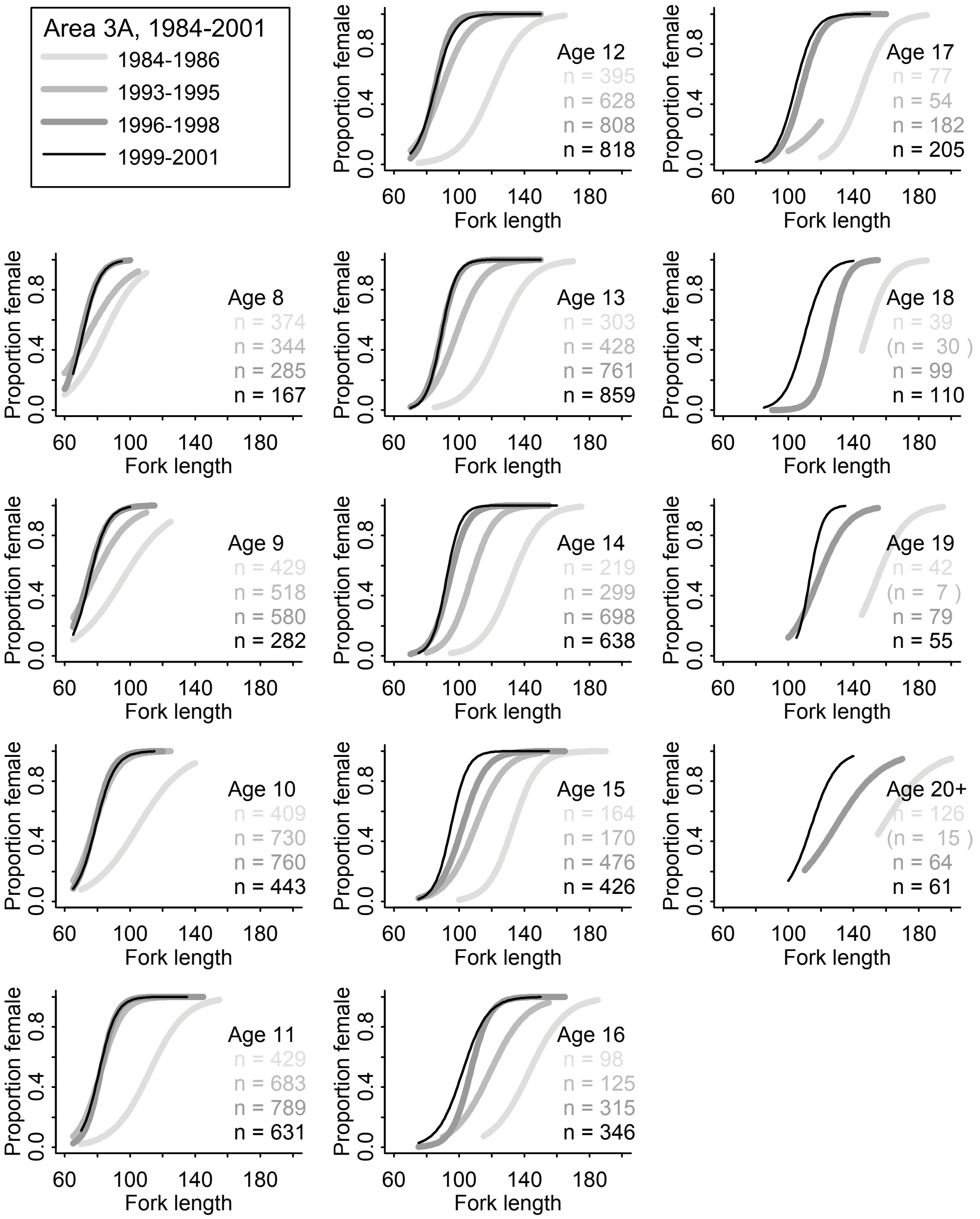


Figure B6b. Logistics fitted to proportion female at length in Area 3A by three-year periods.

There is a good deal of year-to-year variation in the sex ratio at a given age, in both the survey observations and the commercial estimates. It is doubtful that the sex ratio at age in the commercial landings really varies much from year to year, so it seems sensible to smooth the commercial sex ratio estimates at each age over years to compute the catch at age by sex in each year. Doing so makes very little practical difference because the year-to-year variability of estimated catch at age by sex is due almost entirely to the year-to-year variability of estimated catch at age rather than estimated sex ratio.

Discussion

The procedure developed above seems to be quite reliable because the sex ratio at length within age has a very simple form that appears to be nearly uniform throughout a regulatory area in any given year. But that is only known to be true for the summer survey period. If there is a seasonal variation outside that period, the commercial landings could have a different sex composition. And even if the commercial fishery encounters the same sex ratio as the survey at a given age and length, the sex composition of the catch could be different if the ratio of female to male catchability is not the same as in the survey, e.g. because of a strong effect of bait type on sex-specific catchability. Differences of this sort are somewhat far-fetched but not impossible. While it seems reasonable and worthwhile for the time being to estimate commercial sex compositions as proposed, it remains desirable to devise a method of determining the sex of fish in the commercial samples.

References

Clark, W. G. 1992. Estimation of halibut body size from otolith size. Int. Pac. Halibut Comm. Sci. Rep. No. 75.

Clark, W. G. 2001. Age and size composition and recent trawl and setline catches in Area 3A. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2000:345-352.

Appendix C. Statistical distribution of IPHC age readings

This paper appeared first as:
Clark, W. G. 2004. Statistical distribution of IPHC age readings. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2003: 99-110.

Abstract

This paper reports estimates of the statistical distributions of surface and break-and-burn readings about the true age of a fish. The variance of single readings of a given otolith is low for both types. Surface readings are increasingly biased downward after age 12, and among older fish they have a large variance due to variation among otoliths of a given age in the number of annuli countable by the surface method. A recipe section at the end contains detailed instructions for smearing age distributions.

Introduction

Age readers strive to follow consistent rules when counting annuli on an otolith, but there is still some judgment involved and consequently some variability both within and among readers in the age assigned to a given otolith. A single age reading can therefore be regarded as a draw from a probability distribution. An estimate of the distribution can be incorporated into the stock assessment model to predict the observed distribution of age readings that would result from an underlying true age composition, and thereby sharpen estimates of abundance and especially year-class strengths.

In concept, the distribution of readings of a single otolith is what would be observed if the same otolith were read many times by experienced readers following the same protocol. The mode of this distribution is by definition the correct age to assign to the otolith according to that protocol, whether or not it is the true age of the fish. For clarity this modal age will be called the "canonical age" of an otolith.

At the International Pacific Halibut Commission (IPHC), otoliths were aged by surface reading until the early 1990s. During the 1990s an increasing number of difficult and older otoliths were broken and burned for reading because it was known that surface readings tended to be too low in those cases, but surface readings were continued for a majority of fish until 2002, when surface reading was discontinued altogether in favor of breaking and burning.

Break-and-burn readings have recently been validated by comparing them with a reference chronology of ^{14}C uptake resulting from nuclear tests in the mid-20th century (Piner and Wischniowski in review), so we now know that the break-and-burn protocol is accurate; i.e. that the canonical break-and-burn age of an otolith is the true age of the fish, and that the canonical surface age is too low, at least for older fish. The problem is therefore to estimate the distribution of both surface and break-and-burn readings at each canonical break-and-burn age.

Distribution of break-and-burn readings about the canonical age

Given a sufficiently large sample of paired readings, the distribution of deviations of single readings from the canonical age can be estimated by fitting the sample distribution of differences between paired readings (Clark 2004). This procedure shows that the unsigned deviations of break-and-burn readings follow a geometric distribution. To be specific, let the random variable *b* denote a single reading of an otolith of canonical age *B*, and let $v = b - B$ be the signed deviation, so *v* has the same distribution as *b* except that it is shifted so as to have a mode at zero. Then the probability of observing a given unsigned deviation $|v| = 0,1,2,\ldots$ is $f(|v|) = p \cdot q^{|v|}$ where $q = 1 - p$. The distribution is assumed to be symmetric, so the probability of observing a given *signed* deviation *v* is $f(v = 0) = p$ and $f(v \neq 0) = p \cdot q^{|v|}/2$. The single parameter $p = f(0)$ decreases with increasing age as the variance of readings increases, and it can be computed from the sample variance σ_v^2 of the readings at each age with the formula

$$p = \frac{-3 + \sqrt{1 + 8 \cdot \sigma_v^2}}{2 \cdot \left(\sigma_v^2 - 1\right)}$$

Otoliths can and must be grouped by mean assigned age for this purpose because the canonical age of an otolith is unknown. The variance of the signed deviations σ_v^2 can be estimated conveniently as half the variance of the signed difference between paired readings.

The standard deviation of break-and-burn readings increases with canonical age in a non-linear fashion (Fig. C1), apparently leveling off at some point. The shape of this curve at ages beyond 25 or so is not well determined by the amount of data presently available; it should be re-estimated in the future. For the time being, the trend is well-described by the fitted curve, which is:

$$\sigma_v(B) = 1.28 \cdot \left(1 - \exp\left(-0.100 \cdot (B - 2.93)\right)\right)$$

where *B* is the canonical break-and-burn age (and the true age).

Parenthetically, the unsigned deviations of surface ages also follow a geometric distribution, and the standard deviation of surface age readings increases linearly with canonical *surface* age:

$$\sigma_v(A) = -0.112 + 0.0668 \cdot A$$

where *A* is canonical *surface* age, which for fish older than 12 or so is less than the canonical break-and-burn age. When corrected for the low bias of surface ages among older fish, this equation produces values similar to the standard deviations of break-and-burn readings for fish of the same true age: around 0.5 y at age 10 increasing to about 1 y at age 20 and continuing to increase thereafter. So for both kinds of reading the variance about the respective canonical age is modest.

Mean of surface readings at a given break-and-burn age

Through about age 12 there is little difference on average between surface and break-and-burn readings. (In fact surface ages are on average slightly higher for younger fish, but the difference is negligible.) Beyond a break-and-burn age of 12, however, surface readings are lower on average, by a growing margin (Fig. C2). The relationship between assigned break-and-burn age *b* and mean surface age μ_a is well described by the curve shown in the figure, which is:

$$\mu_a(b) = b \qquad \text{b} \leq 12$$

$$\mu_a(b) = 26.58 \cdot \left(1\text{-}\exp\left(\text{-}0.0614 \cdot (\text{b-}2.17)\right)\right) \qquad \text{b} > 12$$

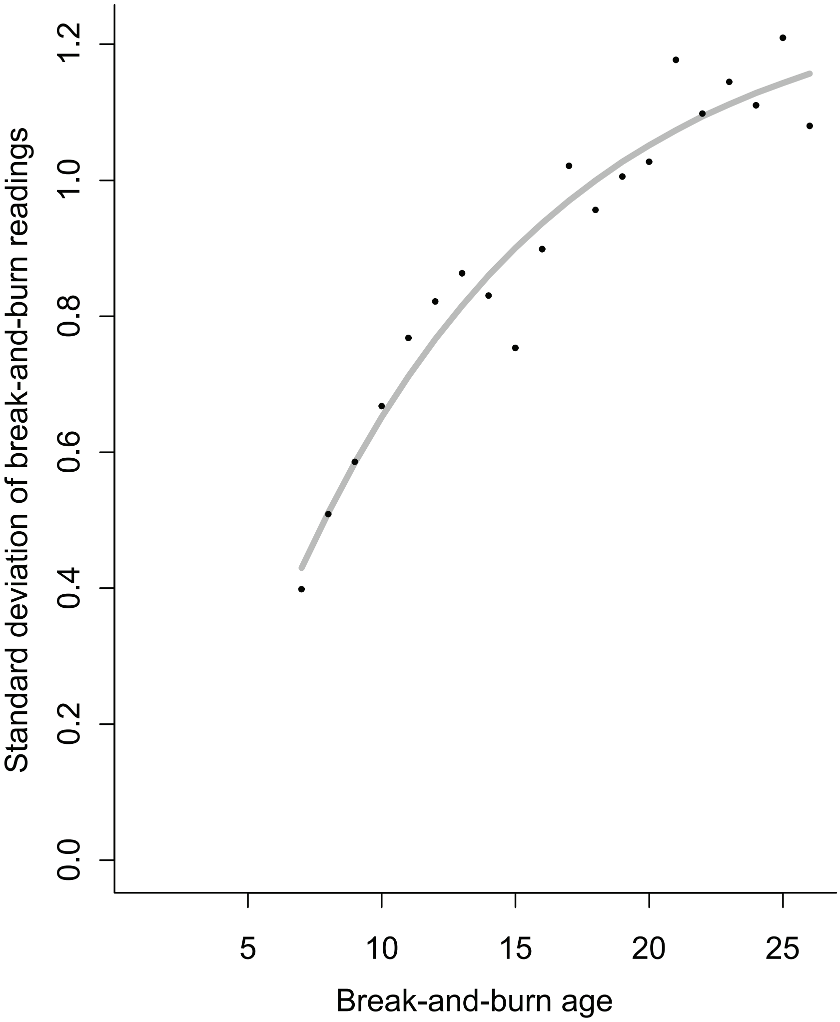


Figure C1. Standard deviation of break-and-burn readings as function of age. The abscissa is the mean assigned age in paired readings. The ordinate is the root of half the variance of the signed difference of paired readings. The gray line is the fitted curve; see text.

This is not precisely what we want, which is the relationship between *canonical* break-and–burn age *B* and mean surface age. The fitted curve is actually an errors-in-variables regression, but simulations show that the estimates are very close to the correct values, mainly because of the low variance of the break-and burn readings.

The data plotted in Figure C2 consist of all paired readings from 1992 through 2002, representing all regulatory areas and both sexes, some read both ways not according to any experimental design but because of difficulty in assigning a surface age. So while large, it is a mixed and not entirely random sample. But the relationship between break-and-burn and surface ages is very similar for subsets of the data grouped by area or sex or reading type, so all of the data were pooled to compute a single working formula.

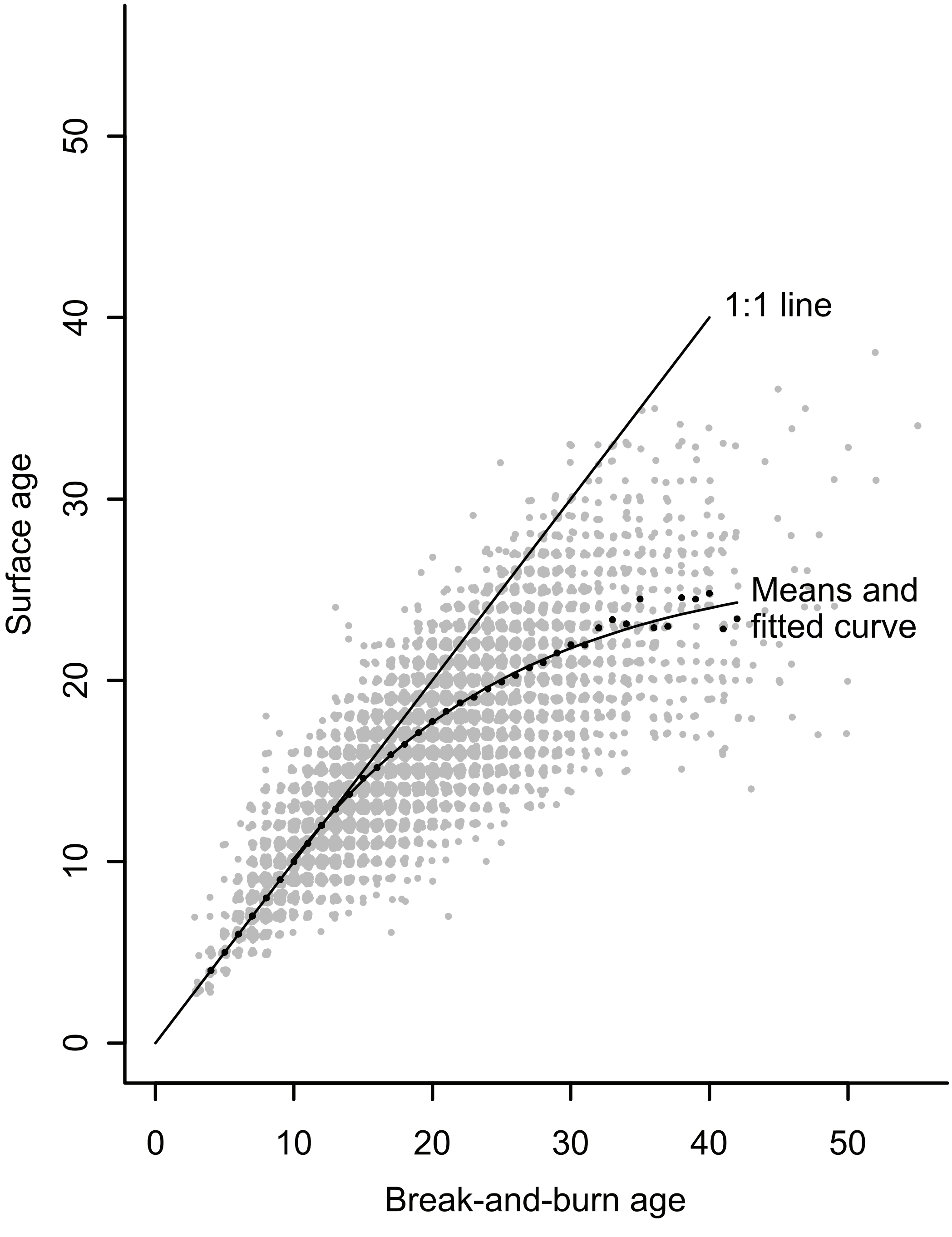


Figure C2. Surface age reading plotted against break-and-burn age reading of the same 60,000 otoliths. The gray masses are the raw data points (jittered). The black points are the mean surface age at each break-and-burn age.

Variance of surface readings at a given break-and-burn age

The relationship between break-and-burn reading and the standard deviation of surface readings and is well described by the fitted logistic shown in Figure C3, which is:

$$\sigma_a(b) = 0.78 + 3.98 \Big/ \left(1+\exp\left(-0.189\cdot(b-24.79)\right)\right)$$

This variance is inflated because the otoliths having a break-and-burn reading of, say, b' in fact consist of a mixture of canonical break-and-burn ages. Let B' denote that mixture, and let *a* and *A* denote assigned and canonical surface ages. The variance of *a* for a given b' can be represented as the sum of contributions from the component ages $B \in B'$ by employing the rule that an unconditional variance is equal to the expectation of the conditional variance (over the component ages) plus the variance of the conditional expectation (over the component ages). Thus:

$$\sigma_a^2(b') = V(a|b') = \underset{B'}{E}\left[V(a|b',B)\right] + \underset{B'}{V}\left[E(a|b',B)\right]$$

For a given otolith the surface and break-and-burn readings are statistically independent, so in the first term in the sum $V(a|b',B) = V(a|B) = V(A|B) + V(a|A)$ and in the second term $E(a|b',B) = E(a|B) = E(A|B)$. Let $m(B) = E(A|B) \approx \mu_a(b')$ represent the relationship described above between canonical break-and-burn age *B* and mean canonical surface age *A*. With these substitutions:

$$\begin{aligned} V(a|b') &= \underset{B'}{E}\left[V(A|B)+V(a|A)\right] + \underset{B'}{V}\left[m(B)\right] \\ &\approx V(A|B=b') + V(a\,|\,A=m(b')) + m'(b')^2\cdot V(b|B=b') \end{aligned}$$

The form of the third term above relies on the fact that the distributions $f(B|b)$ and $f(b|B)$ are approximately equal, so $V(B|b) \approx V(b|B)$. This is just the variance of the break-and-burn readings about the canonical break-and-burn age. Here it is multiplied by the square of the slope of $m(B)$ at b' because locally $m(B) \approx m(b') + m'(b')\cdot(B-b')$ so $V(m(B)) \approx m'(b')^2 \cdot V(B|b')$. Among older fish this slope is low because surface age changes slowly with true age, so the third term contributes little to $V(a|b')$ at those ages.

The second and third terms can be computed from the known variances of surface and break-and-burn readings about their respective canonical ages. The first term is the variance of canonical surface age at a given canonical break-and-burn age. If all otoliths of a given canonical break-and-burn age belonged to a single canonical surface age, or a very narrow range, this term would be small. It is not. Through about age 15, it accounts for half or slightly less of the total variance, and beyond age 15 it increases steeply, by age 25 dwarfing the other variance components (Fig. C4).

Form of the distribution of surface ages at a given break-and-burn age

From about age 20 onward, where the variance and therefore the form of the distribution of surface ages $f(a|b')$ is dominated by $V(A|B)$, the distribution is very well approximated (Fig. C5) by a discrete version of the normal density:

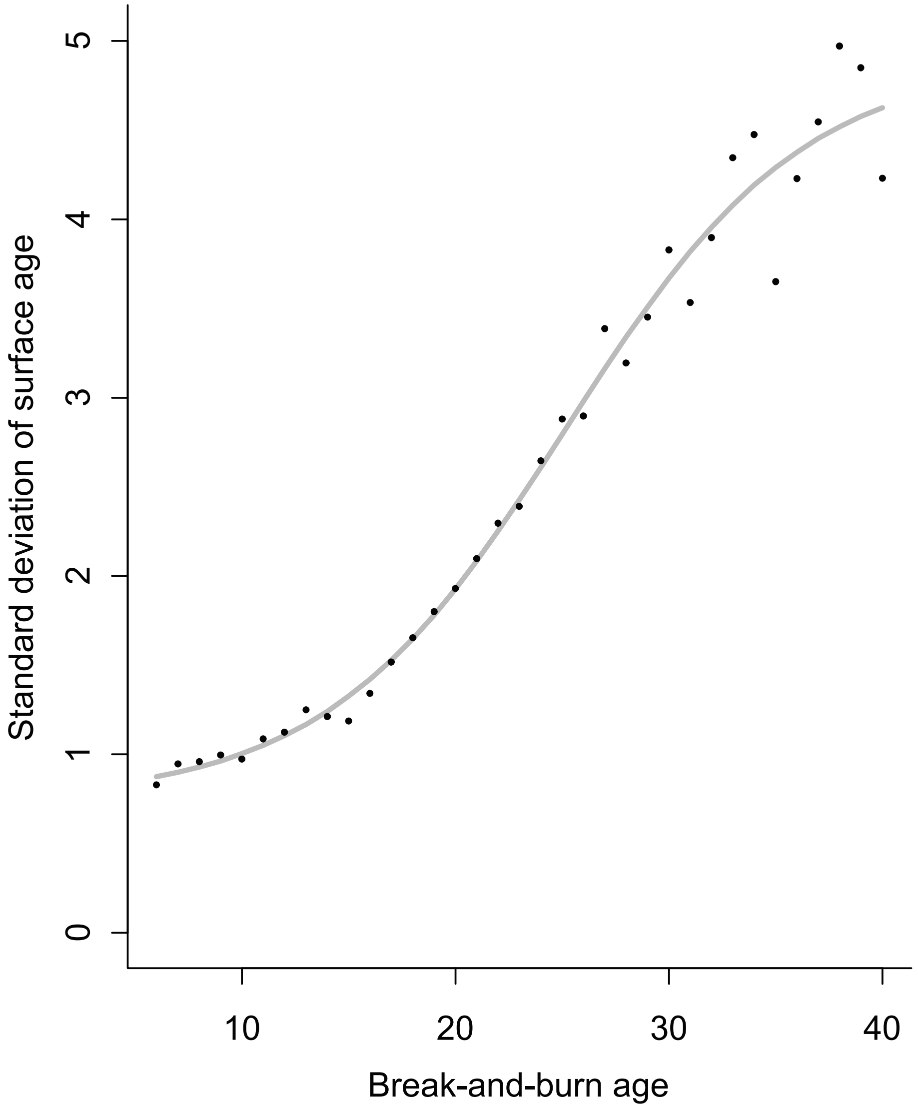


Figure C3. Standard deviation of surface readings plotted against break-and-burn readings (points), and a fitted logistic.

$$f\left(a \middle| b'\right) \propto \exp\left(-\left(a-\mu_a\left(b'\right)^2\right)\Big/\left(2\cdot\sigma_a^2\left(b'\right)\right)\right)$$

The standard deviation of this density is equal to the parameter σ_a for $\sigma_a \geq 0.6$; it is less than σ_a for smaller values.

Among younger fish, where other variance components are significant, the distributions are leptokurtic, having more data points at the mean and in the tails (and fewer in between) than a normal distribution with the same variance. This is clear in Fig. C5, where at low ages the normal distributions with parameters equal to the sample moments (thick gray lines) fail to reach the high observed frequencies of the modal (and true) ages. This shortfall can be prevented by using an ad hoc scaler to reduce the working value of σ_a to something less than the sample standard deviation. Specifically, if the sample standard deviation is multiplied by a scaler that increases linearly from 0.5 at age 5 to 1.0 at age 20, the normal approximation is adequate for the younger ages as well. The black lines in Figure 4 are predicted frequencies based on the discrete normal distributions with means and standard deviations taken from the fitted curves reported above, except that at younger ages σ_a is calculated by scaling down the fitted standard deviation.

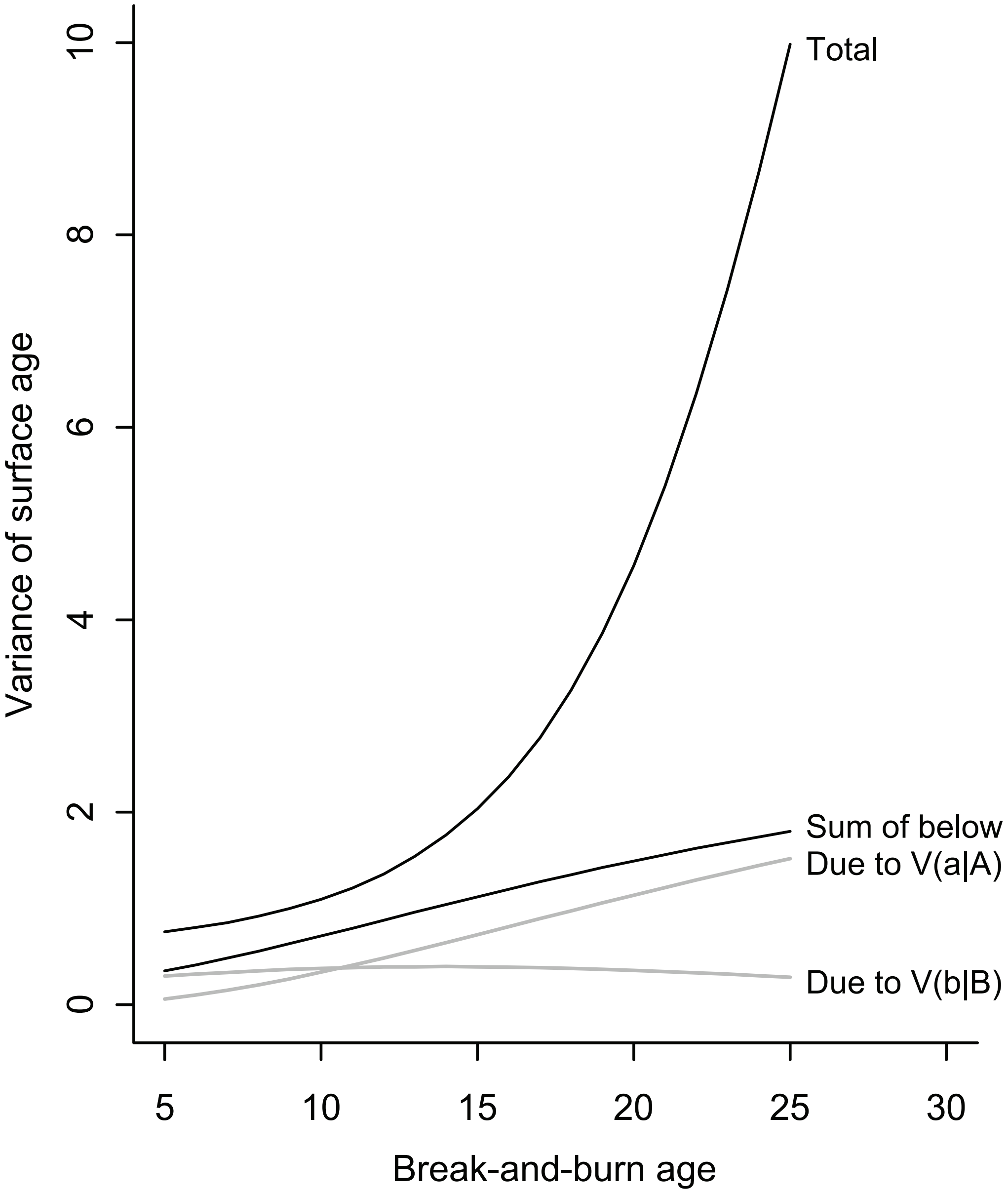


Figure C4. Components of the variance of surface readings at a given break-and-burn reading. V(a|A) is the variance of surface readings about the canonical surface age. V(b|B) is the same for break-and-burn readings, but only a fraction enters the total variance; see text. The difference between the two upper lines is due to V(A|B), the variance of canonical surface age at a given canonical break-and-burn age.

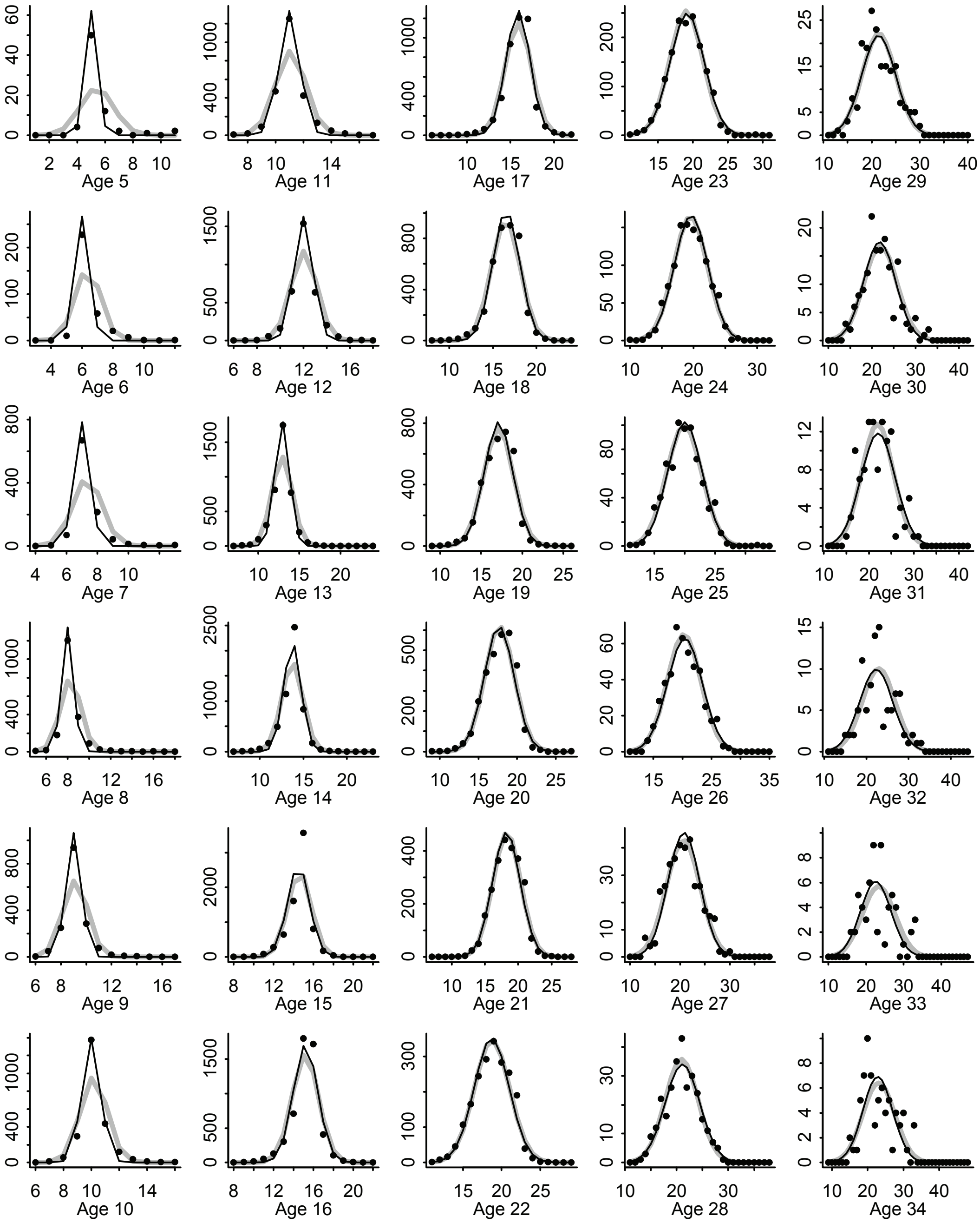


Figure C5. Observed and predicted frequencies (number of otoliths) of surface ages grouped by assigned break-and-burn age. The gray line in each plot is a discrete version of the normal density with parameters equal to the sample moments. The black line is a normal density with parameters calculated from the fitted curves reported in the text.

Discussion and conclusions

Predicting the distribution of break-and-burn readings of fish of a given true age *B* is straightforward because that is also the canonical break-and-burn age, so the readings can be expected to follow the simple geometric distribution described above and detailed in the recipe section at the end.

Surface age readings are more complicated because their distribution depends on both the distribution of individual surface readings about the canonical surface age, $f\left(a\middle|A\right)$, and the distribution of canonical surface age at a given true age or, equivalently, canonical break-and-burn age, $f\left(A\middle|B\right)$. Beyond age 15 or 20, this source of variation is quite large. The canonical surface age of an otolith is the number of surface-countable annuli as defined in the protocol, and age readers can make that determination with a high degree of consistency. But among older otoliths of the same canonical break-and-burn age there is obviously a great deal of variation in the number of surface-countable annuli, which is not surprising.

The distribution $f\left(A\middle|B\right)$ could be estimated by modeling, and surface readings of fish of each true age *B* could then be predicted in two steps by predicting the canonical surface age distribution and then the surface age reading distribution. A simpler alternative is to predict the distribution of break-and-burn readings $f\left(b\middle|B\right)$ for all ages to obtain the overall marginal distribution of predicted break-and-burn readings and then use the simple model of $f\left(a\middle|b\right)$ developed above to predict the corresponding distributions of surface readings. This approach has the attraction that both $f\left(b\middle|B\right)$ and $f\left(a\middle|b\right)$ can be (and have been) estimated directly from the available data. The smearing procedure is spelled out below.

References

Clark, W. G. 2004. Nonparametric estimates of age misclassification from paired readings. Can. J. Fish. Aquat. Sci. 61:1881-1889.

Piner, K. R., and S. Wischniowski. 2004. Description of a Pacific halibut chronology of bomb radiocarbon from 1944-1981 and a validation of ageing methods. J. Fish. Biol. 64:1060-1071.

Recipes

This section is intended as a working reference for predicting the distribution of break-and-burn and surface readings of otoliths of a given true age *B*; it contains no new material.

Distribution of break-and-burn readings

Let $v = b - B$, the signed deviation of a single reading from the true age. The distribution of v is $f(0) = p$ and $f(v \neq 0) = p \cdot q^{|v|}/2$ where $q = 1 - p$ and the parameter $p = f(0)$ is a function of the variance of break-and-burn readings at age *B*:

$$p = \frac{-3 + \sqrt{1 + 8 \cdot \sigma_v^2}}{2 \cdot \left(\sigma_v^2 - 1\right)} \quad \text{if } \sigma_v^2 \neq 1 \text{ else } 2/3$$

$$\sigma_v^2(B) = \left(1.28 \cdot \left(1 - \exp\left(-0.100 \cdot (B - 2.93)\right)\right)\right)^2$$

Distribution of surface readings

The first step is to generate the distribution of predicted break-and-burn readings (assigned ages) *b* for fish of all true ages, including *B*, as described above. For each assigned age b', the mean μ_a of predicted surface readings is:

$$\mu_a(b') = b' \qquad \text{b}' \le 12$$

$$\mu_a(b') = 26.58 \cdot \left(1\text{-exp}\left(\text{-}0.0614 \cdot (\text{b}'\text{-}2.17)\right)\right) \qquad \text{b}' > 12$$

The actual standard deviation of surface readings is:

$$\sigma_a(b') = 0.78 + 3.98 \Big/ \left(1 + \exp\left(-0.189 \cdot (b' - 24.79)\right)\right)$$

but the computations use a working value $\delta_a = c \cdot \sigma_a$ that is scaled down at the lower ages. The value of the scaler *c* is:

$$c = 0.5 \qquad \text{b}' < 5$$

$$c = 0.5 + 0.5 \cdot (b' - 5)/15 \qquad 5 \le \text{b}' \le 20$$

$$c = 1 \qquad \text{b}' > 20$$

Let $d(a) = \exp\left(-(a - \mu_a)^2 \Big/ \left(2 \cdot \delta_a^2\right)\right)$, a discrete form of part of the normal density function. Note the last term is δ_a^2 not just δ_a. Normalizing the values of $d(a)$ finally gives the density of surface readings at break-and-burn age b':

$$f(a) = d(a) \Big/ \sum_a d(a).$$

Calculation of observed age composition from true age composition

Let $\mathbf{N}_b$ be a row vector of numbers at true age *b* (*b=1,…,B*) and let $\mathbf{MM}_{ba}$ be a misclassification matrix in which each element mm_{ba} is $p(a \mid b)$, the probability of assigning (observing) age *a* $(a = 1, \ldots, A))$ given that true age is *b*. So row *b* of $\mathbf{MM}_{ba}$ is the distribution of observed ages *a* at true age *b*. Then the row vector of expected numbers at observed age $\mathbf{N}_a$ is:

$$\mathbf{N}_a = \mathbf{N}_b \cdot \mathbf{MM}_{ba}$$

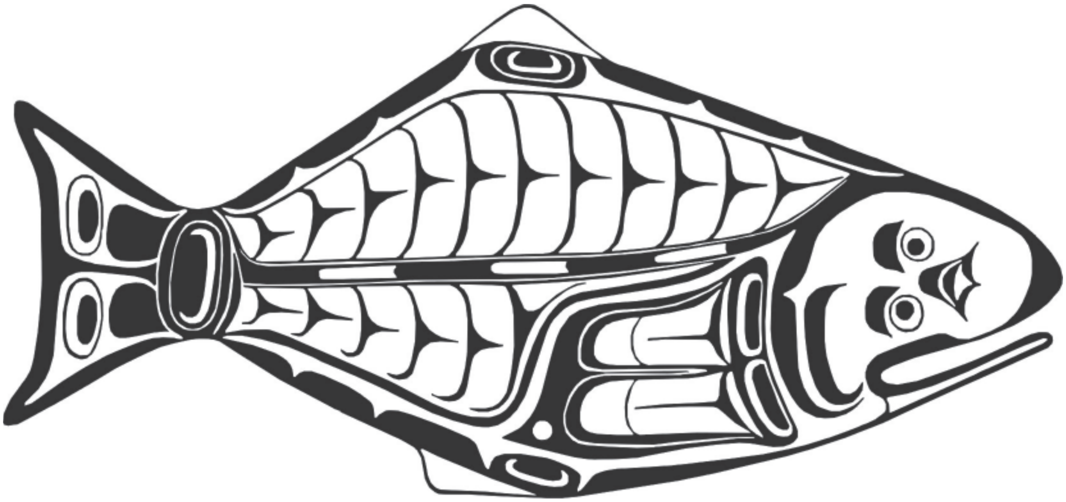
If $\mathbf{N}_{yb}$ is a matrix in which row *y* is a vector of numbers at true age *b* in year *y*, the corresponding matrix of expected numbers at observed age *a* is:

$$\mathbf{N}_{ya} = \mathbf{N}_{yb} \cdot \mathbf{MM}_{ba}$$

Alternatively if the misclassification matrix is set up columnwise, so that column *b* is the distribution of observed ages *a* at true age *b*, then $p(a \mid b) = mm_{ab}$ and the calculation is:

$$\mathbf{N}_{ya} = \left(\mathbf{MM}_{ab} \cdot \mathbf{N}_{yb}^{\mathbf{T}}\right)^{\mathbf{T}}$$

where $\mathbf{T}$ denotes transpose.



Halibut Crest - adapted from designs used by Tlingit, Tsimshian and Haida Indians