Population Assessment

International Pacific Halibut Commission Stock Assessment Workshop

June 27-28, 2007
Nexus Hotel
Seattle, Washington

Note: The following summarizes key points of the workshop and is not intended to be a verbatim transcript. Appended to this summary is a list of questions posed by participants and the staff responses.

Wednesday, June 27

IPHC Director, Dr. Bruce Leaman, introduced the external scientific reviewers Drs. Chris Francis and Paul Medley from the Center for Independent Experts (contracted through University of Miami), who attended the meeting as a component of an IPHC independent assessment review, and Dr. Steve Martell of UBC Fisheries Centre as moderator of the workshop.

Dr. Martell's opening remarks included recognizing that there has been a substantial change in the halibut assessment from a closed area to a coastwide approach. This workshop is being held to look at the technical details of the model, the data going into the model, and the method for apportioning the coastwide biomass into IPHC regulatory area biomass. Material concerning the 2006 assessment was made available on the IPHC website prior to the meeting (http://www.iphc.washington.edu/halcom/newsrel/2007/nr20070509.htm).

Dr. Leaman presented an overview of Pacific halibut management: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0701bml.pdf). Discussion of the presentation included clarification of CEY and harvest rate. There was a comment that it appears IPHC deliberates catch limits privately and that the 2007 results were a surprise to the public at the January Annual Meeting. Dr. Balsiger agreed that perhaps the Commission was not explicit enough at the public session in explaining the rationale for how it arrived at catch limits for 2007 and will try to improve in the future.

Ms. Heather Gilroy presented commercial fishery removals with no discussion following: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0702hg.pdf).

Mr. Gregg Williams presented other removals: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0703gw.pdf) There was discussion of assumed observer coverage in the Bering Sea trawl fleet and whether scientists believed that the rate of observer coverage reflected the actual percentage of the catch. It was clarified that the 30% target coverage for vessels 60-125 feet in length and 100% coverage for vessels greater than 125 feet referred to hauls made and not directed catch. Therefore, the actual catch observed could vary from the haul coverage targets. It was noted that for the stock assessment model, the current year's figures for bycatch mortality are used. For some fisheries that is an estimated number for the entire fishery and for others it is a projection based on partial year's data.

Mr. Claude Dykstra presented setline survey information: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0704cd.pdf). There was discussion of the presence of commercial fishing immediately before the survey fishes on a station and how this event is taken into consideration. The NMFS sablefish survey was given as an example of a survey where commercial vessels are asked to cease fishing prior to survey fishing. The IPHC staff noted that there are many survey vessels fishing at the same time and the logistics of limiting access at these sites as the survey progresses would be prohibitive. Furthermore, the survey should reflect what is happening on the grounds and if commercial fishing lowers CPUE at certain sites, then that is the reality. Also, factoring fishing pressure into the CPUE in some fashion would be difficult.

There was significant discussion of the survey design. Some participants suggested that the systematic sampling design is not unbiased to relative abundance in certain areas. For example, the design results in lower variance in areas where the continental shelf is wider and there are more stations. However, the stations are assigned systematically, so that sampling is in proportion to the amount of bottom area, i.e., the wider the shelf, the greater the number of stations. It was also noted that the survey begins at 25 fathoms and there may be some halibut in more shallow areas. The staff explained that these shallow areas have been fished before but that it is technically difficult to fish the shallows.

A discussion took place of how species composition is estimated in the survey. The staff explained that the systematic 20% hook sampling appears to represent abundance of common species fairly well, but is less precise for less common species. There have been two detailed analyses of 20% vs. 100% sampling and both concluded that the present procedures were unbiased but the precision of estimation decreased with lower occurrence of a species. It was also noted that such results occur for any subsampling scheme and are not unique to the IPHC procedure. A 100% hook count for species composition requires a third sampler and the survey vessels generally carry two samplers. The exception is in British Columbia where the commercial fishers have funded a third sampler to account fully for bycatch.

Following a short break, Dr. Ray Webster presented the PIT tag study results: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0705rw.pdf). Following clarification of points from the presentation, it was suggested that if harvest rates are different among areas, then this will result in a redistribution of biomass relative to proportional harvest. The staff commented that it could be happening for example in Area 2 where the harvest rate appears to be higher.

The fact that fish appear to be migrating from Areas 4A and 3B to Area 3A, but virtually no fish appear to be migrating from Area 2C to Area 3A was noted. The staff acknowledged this also and speculated that density dependence could be altering distribution, i.e., a large biomass in Area 3A is inhibiting recruitment into the area. In addition, migration on the eastern side of the stock appears to be primarily in an eastward and southward direction, where recoveries have been sufficient to establish a trend.

There was discussion of Area 4A in terms of Bering Sea side versus Gulf side. The staff noted that the halibut from the Bering shelf Closed Area likely recruit to the Bering Sea side of Area 4A.

The following suggestions were made to staff:

- to analyze whether the fishing mortality is different for the year following release of tags.

- to analyze how the Gulf side compares to the Bering Sea side of Area 4A, and also look at how many halibut actively move from the Bering Sea side to the Gulf side.

Dr. Clark presented IPHC data pre-processing practices: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0706bc.pdf). It was noted that processes of smoothing the data are different for survey versus commercial data. The staff agreed to revisit the topic later in the workshop. There was some discussion of sex composition and the high site fidelity of PIT tagged fish. Dr. Clark noted a study done in 2005 that looked at commercial fishing recoveries within 10 miles of PIT tag stations, and suggested that perhaps with the coastwide model, the proximity to stations should be adjusted. Also mentioned was that mean depth distribution changes from east to west but the model includes 0-300 fm depth in all areas, however staff noted that catch of halibut below 300 fm was very low (< 3%) in all areas.

The following were requests of staff:

- to look at the halibut tagged on the Gulf side of Area 4A and the resulting recovered area.
- to check depth distribution by area and resulting tag recoveries.

Following a break for lunch, Dr. Clark presented the IPHC stock assessment model: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0707bc.pdf). He began with a description of the basic 'vanilla' model. There was significant discussion of catchability. It was explained that the definition of catchability is the proportion of the population that will be caught with one unit of effort. In the assessment, there are different catchabilities assigned to each fishery and in the coastwide version, there is one estimated value each for males and females. Dr. Francis suggested that while using one value for the coastwide catchability for the commercial fishery is likely not a problem, it may be for the survey and that value should be allowed to fluctuate. Staff commented that commercial catchability is allowed to fluctuate in the model but survey catchability is not.

There was some discussion of natural mortality and removal accounts. It was noted that the total groundfish fishery in Area 2B is a different fishery currently than in the past, with 100% of removals in all fisheries now being accounted for. The staff agreed that the better estimates could impact results in the model and that discard mortality numbers in B.C. are much lower than in the past. Ultimately, the most influential variables in the models are the selectivities and the catchabilities.

Dr. Clark continued his presentation with fitting of the model. He noted that the model fits well with data for both females and males in the survey but are divergent with males in the commercial fishery. Several models have been tried and thus far, there are no better fits that cure the retrospective problems. Dr. Francis suggested that instead of fixing the data outside the model, do it inside instead. In addition, fit the CPUE without a scalar at all and perhaps maybe slightly larger ones on other parameters.

Dr. Clark presented penalties and weights with little discussion.

Following a break, Dr. Clark presented alternative model fits. Following some discussion of the data being used for the alternative model fits, the staff agreed to look into reasons for the difference in male and female selectivities in the Freeform and the Smooth Ten models.

Dr. Martell reminded the attendees that the staff were trying to be open about the subjectivity that goes into the model and one purpose for this workshop is to air out that subjectivity and see how different assumptions can lead to divergent views on management outcomes.

There was some discussion of the IPHC staff's evaluation of gains and losses in the different model fits. Dr. Leaman explained that the staff does not expect to go back to closed-area assessments given the compelling evidence of migration and its estimated impacts. The IPHC Commissioners in attendance reserved judgment on this change pending this workshop and further study.

There was further discussion of closed area versus coastwide assessment models. It was pointed out that these two options were at the extremes and that there were other options in between. The staff noted that there are two separate topics, one is the assessment and the other is apportionment. Exploitation is higher in the east than in the west and the coastwide model protects the stock as a whole. There was a suggestion that PAT tags could be used to look at migration rates. The staff responded that the project is possible, but would be costly, and they had been unsuccessful at getting funding for a large PAT tag project. Dr. Francis suggested including information on migratory movements that we already have as well as what-if scenarios into the coastwide assessment model to see how the results change. The staff commented that they are confident that area assessments would be very sensitive to that information, that it is incomplete or lacking in several areas, and its use would encumber the assessment with the same problems as exist in the closed-area assessments. It was agreed that such a process could be a valuable simulation exercise but in the absence of accurate and detailed migration estimates of high precision, the model results would be largely driven by values in which we cannot place great confidence. There would also be a need to reconcile this information with the survey results. It was suggested to treat them as random effects. There was a comment that in the closed-area assessments there were declining trends in the stock that the model did not pick up, and that a better understanding of the reasons for this is crucial. A counterpoint was made that it is not a biological problem, but may be from the timing of how the fishery is executed and results may carry over into selectivity.

Dr. Martell redirected the discussion back to the alternate models. He commented that there is too much weight (and double fitting of age composition data) being put on age and he suggested downgrading the age composition and focusing further on the survey CPUE. The large effective sample sizes on the age-composition information may be one source of the retrospective bias in the assessment model. Age composition alone does not give information on absolute abundance and that is important to remember when making allocations.

Dr. Clark presented area apportionment strategies:
(http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0708bc.pdf). He noted that the declines in survey biomass indices for Areas 3B and 4 were expected given the low historical exploitation in these areas and the 'fishing down' effect of higher exploitation rates in more recent years. Concerning relative catchability among areas, he showed that the recovery rate of PIT tags per 10,000 fish scanned was similar in Areas 3A and 3B. If for instance, catchability was actually higher in an area, the expectation would be a higher recovery rate of PIT tags in that area. There was some discussion of the use of trawl data. Dr. Francis asked the staff to produce for the workshop, an estimate of absolute abundance based on trawl surveys.

A discussion of hook competition ensued. Dr. Clark conducted an analysis of hook competition among areas and noted that except for possibly Areas 2A, 4B and 4D, CPUE is consistent among areas on the survey. He examined bait competition and found that the fraction of baits recovered on survey stations is consistent across Areas 2B-4A. It was further explained by staff that some studies suggest a local depletion effect around the gear in some areas and not necessarily a loss of bait scent. It was noted that competitive interference from bycatch or other halibut could be important, but the analysis indicates that it is not a significant factor in the interpretation of survey CPUE. Dr. Richards asked if there are research projects being designed to deal with these questions. Dr. Clark concluded that setline survey CPUE appears to be a consistent index of density in Areas

2B-4A, and a case could be made for scaling upward in Areas 2A and 4B but there is no objective means to choose the appropriate scalars.

Discussion was opened to look at methods of biomass allocation among areas. It was noted that fundamentally, it is a policy decision, but that decision should be based on sound science and sustainability. Sablefish apportionment was described and it was noted that it is the goal to harvest sablefish at equal rates across the range. From a migratory standpoint, 30-40% of sablefish can move in a year and smaller individuals are more likely to migrate than larger ones.

The final discussion of the day was a summary of apportionments done by the Council for other species such as rockfish, pollock, and cod.

Thursday, June 28th

Dr. Clark presented the results of the alternative model fits that had been suggested in the last session: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0709bc.pdf). The major features changed were to remove any double fitting of data (catch at age but not CPUE at age; male and female catch at age but not total catch at age; total CPUE in number but not total CPUE in weight) and survey CPUE variance scalar set to one. In general, these model modifications had very little effect on either the fit of the model to the data or the resultant estimates of exploitable biomass. Discussion points included:

- the merits of allowing catchabilities for both commercial and survey CPUE to fluctuate each year and how that is distinguished from the error for each year. No conclusion was drawn and the discussion was temporarily tabled.
- the robustness of the estimation for catchability. It was noted that good aging data and a good handle on trends helps.
- the contradictory conclusions of different data sources. One trend suggests that the stock size is decreasing and the other suggests a downward trend only if selectivity has remained constant in recent years. Therefore, the two conclusions might be that there truly is a downward trend, but the other explanation might be that there are more smaller fish and a lower age composition. It was suggested that a third data source, the NMFS trawl survey in this case, may be looked at for information on incoming year classes. Staff noted that exceptionally strong cohorts are generally observed in the trawl surveys several years before their appearance in the exploitable stock. However, it was also noted that cohorts observed in Bering Sea trawl surveys may not index cohorts in the Gulf of Alaska.

Dr. Steven Hare presented the IPHC harvest policy: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0710sh.pdf). He concluded that it appears Area 2 is currently harvested too high and the other areas about normal. It was requested that Dr. Hare look at catchability and harvest levels at the edges of areas. Substantial discussion took place regarding coastwide versus closed-area approach at assessment. It was iterated that there are two components, the assessment itself and then the apportionment. The justification for the Commission's decision to not adopt the coastwide assessment at the 2007 Annual Meeting was made. Dr. Balsiger pointed out that the Commission was not rejecting the coastwide approach, but rather wanted to understand it better. It was further clarified that the coastwide model along with the 20% harvest rate were all part of the same package at the Annual Meeting, and when the

Commission voted not to adopt the coastwide model, they by default agreed to go status quo on the harvest rates as well.

The 60% U.S./40% Canadian split policy for Area 2 catch was discussed. It was noted that the 1979 Protocol to the Halibut Convention between Canada and the United States allowed the Commission after 1981to alter that policy in light of 'pertinent information', which includes estimates of biomass distribution and available yield.

Further discussion points included:

- details of the slow up, fast down policy adopted by the Commission.
- the target versus actual harvest rate in different areas. The goal whether using the coastwide or any assessment model is to harvest all areas at the same rate. Given the survey-based apportionment, that should allow a build-up of biomass in the east where the estimated harvest rate has been substantially higher for the past several years.
- the even application of a harvest policy across all areas and transparency of management.
- the fact that, if the survey-based apportionment is correct, the high exploitation rates on the eastern side of the stock have been partially offset by migration and that is why they may have sustained higher harvest levels. However, there has been an increase in exploitation in the western areas since 1998 and the absolute number of migrants may now be lower than historically.

Dr. Hare presented how the IPHC treats bycatch and sport catch data sets in the assessment: (http://www.iphc.washington.edu/halcom/meetings/workshop2007/presentations/ws0711sh.pdf). Following a break, there was discussion of the data sets, clarification of bycatch impacts, and the types of information available. The areas of concern were limited observer coverage requirements for some fisheries in Alaska, the lack of length data for the sport fishery, and the possibility of visiting the impact of the sublegal mortality on the sector of the fishery from which it came. The issue concerning vessels in Alaska being able to choose when to take an observer and the resulting assumed observation rate was also raised.

Dr. Martell summarized the proceedings with three points:

1. There has been a radical change from the closed area to a coastwide assessment. The closed area assessment had problems with sparsity of data, conflicting data sets, and the assumption of closed populations. The rationale for adopting a coastwide approach was to avoid these problems. However, the change introduces additional assumptions; the data are aggregated and the way they are analyzed is not insensitive to potential differences in catchability in each area.
2. Regarding apportionments, right now the setline survey is used with an assumption of constant catchability among areas. Bathymetric contours are different and if the area habitats were mapped, there would likely not be good correlation of catch by depth ranges among areas. This problem needs to be groundtruthed and tagging may be a way to do that.
3. The problem now is what to do in the interim.

The floor was opened to discussion of point 3.

Dr. Jim Ianelli commented that given what was presented for migration rates, the issue of allocation does not have a strong biological basis. Many issues are outside of conservation or scientific concern. He advised that the Commission would be prudent to come up with a formula of constant allocation.

There was a recommendation that the staff blend the commercial and survey CPUE at different rates to see what happens.

Dr. Richards thanked the commission staff and the contributors to the meeting. She noted that the Commission's task will be to make policy decisions around the scientific advice. There should also be some alternate methods of apportionment explored instead of CPUE only. Forecasting is still not comfortable and looking at longer term shifts in effort and the resulting effects to achieve an optimally harvested stock, needs more work.

Dr. Balsiger agreed that the workshop was helpful in his gaining understanding of the model.

Dr. Leaman noted that there are two processes; this workshop and then the independent review. A report of the proceedings and conclusions will be worked up by staff and available around the beginning of August.

Comments on the structure and content on this and of future assessment workshops included:

- this workshop was valuable for those not trained as scientists and/or in stock assessment to better understand the process, but peer reviews (e.g. STAR panels) are also encouraged.
- Area 2A representatives suggested a meeting with staff within the next couple of months to further discuss the concepts. Area 2B representatives were also invited.
- a peer review every few years to devise pro-rated apportionments instead of having the apportionment fluctuate with survey CPUE annually.
- a recommendation for less modeling and more groundtruthing.
- a recommendation that while this meeting, with the hybrid of a peer review and industry workshop was very helpful, perhaps every other year would be enough for the future.
- urging the IPHC to continue with this type of forum and taking impacts of apportionment decisions on small communities such as in Area 2A, into consideration.
- a recommendation that all user groups including charter fleet participate in this type of process.
- a recommendation for future workshops to focus on only one or two aspects of the assessment such as migration, or the model, etc.

Dr. Leaman agreed to make the presentations available via the IPHC website and thanked the attendees and staff for their participation.

Meeting adjourned.

Attendees

IPHC staff

Bill Clark
Claude Dykstra
Heather Gilroy
Steven Hare
Tom Kong
Bruce Leaman
Tim Loher
Lauri Sadorus
Ray Webster
Gregg Williams

IPHC Commissioners

James Balsiger
Ralph Hoard
Laura Richards
Gary Robinson

Invited participants

Steve Martell – Convener, UBC Fisheries Centre
Chris Francis – CIE External Peer Reviewer
Paul Medley – CIE External Peer Reviewer

Others

First Name	Last Name	Title	Organization/Agency
Bob	Alverson	Executive Director	Fishing Vessel Owners Assoc
Kerim	Aydin	Supr. Fishery Biologist	NMFS/AFSC
Ashleen	Benson	Ph.D. Student	Simon Fraser University
Don	Bodenmiller	Fish Biologist	ODFW
Terri	Bonnett	Halibut Coordinator	Fisheries and Oceans Canada
Dave	Carlile	Fisheries Scientist	ADF&G
Tom	Casey	Consultant	
Sean	Cox	Professor	Simon Fraser University
Nick	Delaney	Comm. Fisherman	Kodiak Vessel Owners Assoc
Yvonne	deReynier	Groundfish Branch Chief	NMFS/NWR
Jane	DiCosimo	Staff specialist	NPFMC
Martin	Dorn	Fishery Biologist	NMFS/AFSC
Sharron	Elwood	Comm. Fisherman	
Garrett	Elwood	Comm. Fisherman	
Wes	Erikson	Comm. Fisherman	HAB
Yongwen	Gao	Researcher	Makah Fisheries Management
Dana	Hanselman	Fishery Biologist	NOAA
Thomas	Helser	Fishery Biologist	
James	Ianelli	Assessment scientist	NMFS/AFSC
Tom	Jagielo	Senior Research Scientist	WDFW
Steve	Joner	Fishery Biologist	Makah Fisheries Management
Robert	Jones	Marine Biologist	NW Indian Fisheries Commission
Jacquelynne	King	Scientific Advisor	Fisheries and Oceans Canada
Linda	Kozak	Consultant	Kodiak Vessel Owners Assoc
Loh-Lee	Low	Scientific Advisor	NOAA
Sandra	Lowe	Research Fisheries Biologist	NMFS/AFSC
Joe	Macinko	Comm. Fisherman	
Charles	McCallum	Fishery Biologist	Lake and Penninsula Borough
Scott	Meyer	Fishery Biologist	ADF&G
John	Moller	Fleet Manager	Adak Fisheries
Melvin	Moon	QNR Director	Quileute National Resources
Kris	Northcut	Harvest Manager	Quileute National Resources
Peggy	Parker	Executive Director	HANA
Stan	Sargent		
Joe	Schumacker	Fishery Biologist	Quinault Indian Nation
John	Secord	Comm. Fisherman	Halibut Advisory Board
Paul	Spencer	Research Fisheries Biologist	NOAA
William	Stockhausen	Fishery Biologist	NOAA
Russell	Svec	Fishery Manager	Makah Fisheries Management
Chuck	Tracy	Staff Officer	PFMC
Theresa	Tsou	Senior Research Scientist	WDFW
Jack	Turnock	Assessment scientist	AFSC
Shizhen	Wang	Fisheries Biometrician	Quinault Indian Nation
Tom	Wilderbuer	Flatfish assessment scientist	NOAA
Mark	Wilkins	Survey manager	AFSC/RACE
Gary	Williamson	Comm. Fisherman	Delta, BC
Robert	Wurm	Comm. Fisherman	Linden, WA
Phillip	Wyman	Comm. Fisherman	AK Longline Fisherman’s Assoc.

Appendix
Questions from the stock assessment workshop, June 2007

IPHC Staff

1. **What portion of mark recoveries from Area 4 were released north rather than south of the Aleutian chain?**

The following table shows all commercial 2003-2006 tag-recoveries from Area 4 releases by recovery area. Here Areas 4A and 4B are divided into their northern (Bering Sea: 4A.bs, 4B.bs) and southern components (Gulf of Alaska: 4A.goa, 4B.goa). The dashed horizontal and vertical lines separating the Bering Sea and Gulf of Alaska areas are to facilitate comparison of movement within and between the two large geographical regions.

	Recovery area									
Release area	**2B**	**2C**	**3A**	**3B**	**4A.goa**	**4B.goa**	**4A.bs**	**4B.bs**	**4C**	**4D**
4A.goa	8	7	11	6	12	1	1	0	0	0
4B.goa	1	1	0	0	0	4	0	0	0	0
4A.bs	1	0	2	0	1	1	14	1	1	0
4B.bs	0	0	1	0	0	0	0	1	0	0
4D	1	0	1	0	0	0	1	0	0	15

		Recovery area									
Release		bs → bs				bs → goa					
area	**No. rel.**	**4D**	**4C**	**4B.bs**	**4A.bs**	**4B.goa**	**4A.goa**	**3B**	**3A**	**2C**	**2B**
4D	979	15	0	0	1	0	0	0	1	0	1
4B.bs	347	0	0	1	0	0	0	0	1	0	0
4A.bs	1285	0	1	1	14	1	1	0	2	0	1
4B.goa	789	0	0	0	0	4	0	0	0	1	1
4A.goa	2171	0	0	0	1	1	12	6	11	7	8
		goa → bs				goa → goa					

Only one fish released in the Gulf side of Area 4 was recovered in the Bering Sea, a 4A fish that stayed in 4A. Five (out of 21 recovered) 4A Bering Sea releases crossed to the other side of the islands. One (of two) 4B Bering Sea fish moved out of the Bering Sea (showing up over in 3A), along with two out of 18 Area 4D fish (one went to Area 3A and one to Area 2B). These numbers do not account for recovery rates or scanning rates, but they appear to indicate that the majority of the fish tagged and released in the Bering Sea has stayed in the Bering Sea, although there is clearly enough transfer to avoid any genetic segregation of Bering Sea and Gulf of Alaska fish.

2. Does depth distribution of halibut vary from east to west?

Figure 1, shown at the workshop, shows the relationship between depth and survey CPUE in each area for the years 2001-2006. The plots show the depth effect from a generalized additive model fit in which year was a factor and depth was a smooth term. Using peak CPUE as a measure of distribution, halibut are distributed with peak abundance occurring between about 75-150 in the eastern portion of the stock but extending down to about 200 fm for the western areas (Area 3B and westward). However, in all areas the range of significant halibut abundance, as well as catch by the commercial fleet, is covered by the distribution of survey stations from 25-275 fm.

3. Does the estimated sampling variance of the commercial catch at age/sex include the variance of the proportion female estimated with the fitted logistics?

Yes; the standard multinomial variance based on sample size is scaled up by 1.05^2 to incorporate a 5% coefficient of variation of the estimated proportion female.

4. What does the site fidelity seen in survey recoveries of PIT tags mean for the mark-recapture analysis?

Last year 83 PIT tagged fish were recovered on the setline survey, with 66 of these (around 80%) recovered on the survey station on which they were released. These raw recovery data imply a high degree of site-fidelity of tagged fish, and potentially that tagged fish do not mix well with the untagged population, at least during the summer months when the survey fishing (and tagging) occurs. As almost all recoveries come from commercial fishing, poor mixing will be a problem if there is a mismatch between tag-release locations and commercial fishing locations and this segregation persists throughout the year, even if the fish move off the survey locations. If this is true, on average, an individual tagged fish will be less likely to be recovered than an untagged fish, and the resulting low recovery rates will lead to negatively biased estimates of rates of commercial fishing mortality from the tag-recovery modeling. Preliminary analysis shows some evidence for a mismatch of the distribution of commercial catch location and tag-releases. However, we note that the setline survey recoveries may be misleading, in that fish that are not on or near a station at the time of the survey are less likely to be recovered, and so the true degree of site fidelity is likely to be overestimated from the raw survey data. More detailed examination of these data and the effect of poor mixing of tagged and untagged populations are ongoing.

5. Should the weighting of survey data in the model be time and area invariant or should there be both temporally- and spatially-dependent weighting terms?

This question may have arisen out of some misunderstanding. The coastwide survey CPUE is presently calculated by weighting the area-specific CPUE by bottom area, and calculating a CPUE for each year independently, so this is already being done.

6. Why does the survey apportionment procedure use a 3-year running mean of survey CPUE rather than e.g. a 5-year forward-weighted average?

We have used an unweighted 3-year running mean for doing survey apportionment among several IPHC regulatory areas since the mid-1990s. Adding years makes the running mean more susceptible to bias resulting from trends in the CPUE series if the trends among areas are different, which they clearly are. Forward weighting increases the variance of the running mean. We conducted trials with various options and found the 3-year running mean performed well in terms of bias and variance for trends up to 10% per year. It is stable in practice, too. The estimated area apportionments for the last ten years (Fig. 2) change over time but in a gradual fashion.

7. Can NMFS trawl survey data be used to check for differences in setline survey catchability among areas? How do you know that trawl survey catchability doesn't vary among areas, too?

Trawl survey catchability of halibut definitely varies among areas. It is clearly low in areas with a lot of rough bottom, like 2C and 4B. But between Areas 3A and 3B, where there is a broad, mostly trawlable shelf, there seems little reason to suspect a trawl survey catchability difference. The lack of coincident trawl and setline surveys, conducted with the same fishing gear in all areas, precludes a stock-wide comparison.

8. Do we have enough observations of incoming cohorts to try to estimate them? Are noisy observations of these age groups causing the retrospective pattern?

The first few estimates of incoming cohorts are always noisy, but that in itself should not produce a retrospective pattern.

9. Can the model reliably distinguish between strong incoming recruitment and high mortality of fully recruited fish?

Yes; the stock assessment estimates total mortality by tracking individual cohorts through the fishery, not by examining the age composition in a single year.

10. Some data are fitted more than once, e.g. the survey catch at age/sex and the survey CPUE at age/sex and the total survey CPUE? Isn't that redundant?

Yes, it is. In principle the catch at age/sex and the total CPUE contain all the information, and the CPUE at age/sex could be left out of the fit. Or the model could be fitted just to the survey CPUE at age/sex. But those non-redundant fits are different from each other because the data contain variance. Fitting to all three datasets requires the fit to track cohorts as well as fitting the catch at age and the total CPUE, which are all good things.

At the workshop the model was fitted with no redundancy and there was no effect on the biomass estimates.

11. Were the variance scalars estimated with raw rather than robust deviations?

Yes. In the coastwide data set there are very few outliers so as a practical matter this is not an important issue.

12. Were variance scalars estimated for all data types or only for catch at age/sex?

For all data types. The table below, shown at the workshop, has the working values of tau for each data type. The variance scalar is tau squared. The values of tau are mostly 2-3, meaning that sampling variance accounts for a quarter or less of the total variance of the observations about the model predictions. The remainder is process error and model specification error. The generally good fits of the parsimonious production model indicate that the model structure is appropriate, so most of the variance is process error.

	Females	Males	Total
Commercial catch at age	3.0	2.4	3.5
Commercial CPUE at age	3.1	2.6	3.7
Commercial total CPUE in number			1.6
Commercial total CPUE in weight			1.9
Survey age composition	1.3	1.3	1.6
Survey CPUE at age	2.5	2.6	2.9
Survey total CPUE in number			2.6
Survey total CPUE in weight			2.3

13. Are the biomass outputs calculated using the estimated true age composition and the estimated size at true age?

No. Those calculations use the observed size at the observed age. So, for example in a year with surface age readings, the spawning population in number at true age is calculated and then the corresponding surface age distribution is predicted by smearing the ages, and those numbers at each surface age are multiplied by the observed weight at each surface age to calculate the spawning biomass.

14. The survey apportionment of the estimated coastwide biomass among areas depends critically on survey catchability being equal in all areas. How can that be ground-truthed?

Lacking some independent, indisputable measure of relative density, it is impossible to know whether survey catchability is the same in all areas. It certainly varies among years, so it may well vary among areas within years. Because the survey apportionment is based on a 3-year running mean CPUE, the important question is whether the average survey catchability varies among areas. At the workshop the staff presented some analysis that at least failed to show any difference among areas:

(i) Trawl and setline survey data, although variable, produce similar estimates of relative abundance in Areas 3A, 3B, and 4A where the bottom is mostly trawlable.

(ii) The incidence of PIT tags in commercial landings (tags/10,000 fish scanned) is very similar in Areas 2B, 3A, and 3B, although higher in Area 2C. One would expect a lower incidence of PIT tag recoveries in areas where the survey catchability is lower, because a lower proportion of the stock would have been marked and released on the survey.

(iii) Analysis of survey hook occupancy data indicates that the competition by other species for baited hooks is similar in Areas 2B, 2C, 3A, 3B, and 4A (higher in 2A and lower in 4B and 4D).

15. When the model is fitted, equal weight is given to the catch at age and CPUE data. Why not put more weight on the survey CPUE when that is regarded as the most reliable indicator of stock trends?

Survey and commercial CPUE were given extra weight in some previous assessments, including the closed-area fits that were standard though 2006, because the fits were poor in some areas and extra weight was needed to achieve a satisfactory degree of agreement with the CPUE data. The coastwide fit agrees well with the CPUE series with no extra weight. It is generally good practice to avoid ad hoc weighting, so in the production model all of the weights were set to one. At the workshop several models were fitted in which the weight on the survey CPUE was increased by a factor of 4. The fits to the survey CPUE were nearly indistinguishable from the production model, and the biomass estimates were almost the same. The conclusion was that the unweighted model fits the CPUE data very well.

16. Survey apportionment is based on bottom area from 0 to 300 fm, but the survey only covers 20-275 fm. Would the proportions be different if the survey depth range were used?

New (and old) estimates of bottom area (in square nautical miles) between specified depth contours for all IPHC regulatory areas are presented in the following table. For each area, the percent of the coastwide total is also given. The Area 4A-BS (for Bering Sea) and 4D edge estimates both use 75 fathoms as the shallow contour (instead of 0 or 20 fathoms). The details of the computations are given in a report to be published in the 2008 RARA. A draft of the report may be obtained at http://www.iphc.washington.edu/staff/hare/html/papers/bottom_area.doc. Briefly, the new estimates are derived from a high resolution digital bathymetry dataset while the old estimates derived from hand tracing on NOAA charts.

	20-275 fathoms		0-300 fathoms		0-500 fathoms		0-300 fm. (old)	
Reg. area	N mi^2	%	N mi^2	%	N mi^2	%	N mi^2	%
2A	10561	4.1	13117	4.0	15304	4.4	12000	4.1
2B	22552	8.8	31695	9.7	33237	9.5	28000	9.5
2C	10064	3.9	16316	5.0	17137	4.9	15000	5.1
3A	42871	16.7	50872	15.5	52550	15.1	50000	16.8
3B	23735	9.3	30621	9.3	32289	9.3	30000	10.3
4A-GOA	8929	3.5	10914	3.3	12961	3.7	19000	6.3
4A-BS	7417	2.9	7736	2.4	9057	2.6		
4B	11892	4.6	15411	4.7	23286	6.7	16000	5.5
4D edge	12162	4.7	12405	3.8	13867	4.0	5000	1.7
4CDE shelf	106030	41.4	138670	42.0	138670	29.8	120000	40.7
Total	256213	100.0	327758	100.0	348359	100.0	295000	100.0

The improved digital bathymetry files used as a basis for the 0-300 fm calculations result in more bottom area in Area 4D (at the expense of the Area 4CDE shelf) but relatively small changes (~1%) for other areas. Using only the 20-275 fm range would result in more substantial changes. While the IPHC survey does not extend into any areas shallow of 20 fm, we know from commercial records that catches in these shallower areas are not insignificant: between 3 and 7% in most areas and around 50% in Areas 4C and 4D, thus it would be prudent to include these depths. Conversely, catches from deeper than 300 fm are small: 1% in areas 2A and 3A, 4% in Area 4A and less than 1% in all other areas.

17. At the workshop, the staff showed the ratio of setline survey to trawl survey CPUE at length and claimed that the data showed no difference among Areas 3A, 3B, and 4A. But there did appear to be some differences.

The data clearly showed that the ratio of setline to trawl CPUE in Area 3B was neither 2-3 times what it was in Area 3A nor was it consistently higher in Area 3B, which were the important points in considering whether the closed-area assessments were credible. Similarly, Area 4A shows higher ratios for smaller fish but the same ratios as Area 3A for larger fish. The staff is doing a more detailed comparison of trawl survey and setline survey data.

18. The staff presented GAM estimates of the relationship between depth and survey CPUE but there were no error bars. How precise are those estimates?

Figure 1 shows the fits replotted with 95% confidence intervals.

19. The depth-stratified mean CPUE is the same as the simple mean in all areas except 2A, but in view of the difference there, is there any reason not to compute a depth-stratified mean in all areas?

Not in principle. In practice there might be some confusion about different CPUE series appearing in different places but it is reasonable to use depth-stratified means to accommodate any depth effect. The data suggest that while the effect may not be large, it is not equivalent among areas (Fig. 1)

20. Will the change from closed-area assessments to a coastwide assessment with survey apportionment have a significant effect on capital values?

The estimates of coastwide abundance from the two procedures are about the same, but survey estimate of biomass in Area 2 is only about 15% of the coastwide total, whereas Area 2 has been receiving about 30% of the coastwide total according to the closed-area assessments. A complete implementation of proportional harvest according to the survey apportionment would therefore reduce the yield associated with Area 2 shares by about half, with yield for shares in Area 3B and 4 increasing in value. However, we also estimate that the use of a constant harvest rate policy in all areas would result in an increase of biomass in the eastern portion of the stock, so that the current decreased proportion of the stock in the eastern portion would be only a transitory effect of a survey-based apportionment. Historically, changes in yield associated with shares do not have a direct relationship with capital value because of the change in ex-vessel price per pound that may accompany any changes in yield per share. Increases or decreases in ex-vessel price per pound associated with supply and demand can act to offset changes in yield per share.

21. What apportionment methods other than the survey method could be used?

The setline survey data are the best information available for estimating the distribution of biomass among areas. Trawl survey data would be a possibility if we had comparable data in all areas, but we do not now and never will, because some areas like 2C and 4B are untrawlable. Commercial CPUE is available for all areas, but the comparison of commercial and survey CPUE shows that commercial catchability varies greatly among areas. Commercial CPUE is ten times survey CPUE in Area 2A, about three times in Area 2B, about the same in Area 2C, and so on. These differences do not result from differences in survey catchability; they result from the fishery targeting good grounds more or less effectively while the survey covers the whole area.

The staff has examined several other methods of apportionment, including the historical recruitment distribution as estimated by the closed-area assessments and historical fishery shares. However, none of these other metrics for apportionment incorporates the objective standardization of the survey metric. Historical recruitment estimates are subject to the same errors resulting from migration as the closed area assessments. Historical fishery shares reflect the distribution of fishing effort and are subject to severe biases resulting from the distribution of fishing effort. Using survey data for apportionment is not perfect, as we have noted, but it represents the most objective measure currently available.

Over the long term, we believe yield should be distributed among areas in proportion to biomass. Proportional harvest is standard practice in fishery management for good reasons. It protects the stock against disproportionate harvest of sensitive sub-components of the stock (e.g. behavioral groupings), about which there may be little or no knowledge, but departures can and do occur. The Commission has temporarily assigned catch limits that resulted in non-target harvest rates when there have been significant changes in either assessment methodologies or harvest policies, as a transition to new harvest regimes. However, it now appears that a disproportionate share of the halibut yield has been taken in Area 2 for some time resulting in very high exploitation rates and lower biomass than would result from harvesting at the target rate.

22. How about estimating biomass distribution using a mixture of survey and commercial CPUE?

The IPHC staff had considerable discussion on this proposal. The strongest objection to using commercial data for apportioning biomass is that the raw data consistently show strong differences in commercial and survey catchabilities among areas. The ratio of the two indices varies from 0.48 – 0.99 among areas and is consistent within areas, over time. Introducing commercial data into the apportionment process will embed these biases. However, it can be argued that incorporating some consideration of the commercial data could offset any temporal bias inherent in the survey data, which are collected over only a short portion of the year in each area. On balance, the strongly biased relationship of commercial and survey data convinces the staff to decline the use of commercial data for this purpose.

23. Will an apportionment of yield based on stock distribution at the time of the survey really achieve proportional harvest when fish migrate before, during, and after the survey?

Yes. The concern here is that if we estimate the correct stock distribution at the beginning of the year and allocate yield accordingly, it will be necessary to fish a lot harder in source areas of migration than in destination areas to catch the quotas, because fish will be leaving the source areas and entering the destination areas during the year. That line of reasoning is correct, but the disparity in fishing mortality rates that would result is small.

With a survey apportionment, we do not estimate stock distribution at the beginning of the year but in the middle of the year. Results of emigration and immigration are therefore reflected in survey CPUE, and the rate of fishing mortality is the same in all areas. A full analysis of these effects is posted at http://www.iphc.washington.edu/halcom/research/sa/papers/proportional.pdf.

24. The last three years of data in both closed-area and coastwide fits show declining CPUE that is not fitted well. What's the problem?

The coastwide fit actually tracks survey and commercial CPUE quite well, including the last three years. It is true, however, that survey catchability declined in 2005 and again in 2006. We can see this in model fits where the rate of fishing mortality in 2006 is fixed at various levels and the corresponding series of survey and commercial catchabilities are estimated year by year. In all cases survey catchability is seen to be quite variable among years and to decline in 2005 and 2006.

25. What is the desired distribution of spawning biomass? Will proportional harvest achieve that distribution, or should it modified in some way?

Absent other compelling information, the desired distribution of spawning biomass would be something akin to its distribution absent fishing. Simulation modeling across a range of fishing and migration rates was conducted and reported in the 2007 RARA. The results showed that proportional harvest, i.e., the same constant harvest in all regulatory areas maintained nearly the same spawning biomass distribution as in the unfished state. The unbalanced harvest rates we now believe to have been in effect for at least the past decade – 50% of the target rate in the western areas and 150-200% in the eastern areas – leads to a substantial change in the distribution of spawning biomass. Specifically, the contribution of the eastern areas to the distribution is greatly decreased. At an annual migration rate of 0.06 and instantaneous fishing mortalities in the range of 0.20-0.30, the contribution of areas 2B and 2C to the spawning biomass change from 44% in an unfished state to an equilibrium value of 23-26% when the above described unbalanced harvest rates are applied.

26. To what extent is migration influenced by fishing? In particular, are we seeing migration from west to east because higher exploitation in the east has reduced densities there and created openings for migrants?

The question of density-dependent exclusions has not been investigated for large-scale population distributions. Certainly, studies of territorial fish in both tropical and temperate climates show dominance-based hierarchies of occupation of prime feeding or breeding habitats. The evidence of site fidelity seen in recoveries of PIT-tagged halibut from survey stations provides the potential for such a spatially-explicit behavioral process in halibut. The ubiquity of competitive exclusion as a biological process in populations suggests that higher densities of halibut repetitively occupying the same spatial niches would result in shifts in recruitment patterns relative to periods of lower population densities. Densities of halibut in the central Gulf of Alaska have been at record levels over the past decade, also evidenced by lower growth rates. In conjunction with higher exploitation rates in the eastern portion of the stock, it is reasonable to expect that migration to this eastern region may be higher than it would be under either conditions of lower density in the central Gulf, or lower exploitation rates in Area 2.

27. The survey apportionment assumes that halibut habitat is the same proportion of total bottom area in all areas. Is that true?

The survey apportionment makes no assumption about halibut habitat, which is not well defined in any case. The only assumption about habitat involved in the apportionment is that the survey samples each habitat in proportion to its presence. Survey stations are distributed uniformly in all areas, so they can be expected to sample different kinds of habitat in proportion to their occurrence in each area. An area consisting entirely of good habitat will produce a high CPUE at all stations and therefore a high average CPUE. An area consisting of half good habitat and half poor habitat will produce a high CPUE at half the stations and a low CPUE at half the stations, so its average CPUE will be much lower than that of the good area. Habitat differences are therefore reflected in survey CPUE.

28. How about developing a model with explicit migration in which all of the area-specific data are fitted with area-specific parameters?

Such a model would obviously be the ideal way to accommodate movement of fish. However, it is critically dependent on precise knowledge of the rates of migration by all sizes of fish, at all times, among all areas. Further, if there were any temporal or biomass dependence in such rates, they would have to be estimated continuously. This would be a very large project that would present a number of significant technical difficulties, but the main drawback is that we would not be able to estimate the migration rates internally and the results would depend entirely on what rates we assigned externally. Given the evident difficulties in generating reliable estimates for all sizes of fish, it is highly unlikely that these rates could be known with precision sufficient for making catch limit recommendations.

29. The Commission should recognize that allocation is not a purely biological issue and deal with it by developing an allocation framework that considers both biological and policy issues.

The Commission does recognize that allocation is a subsequent process to biomass estimation. It has traditionally based catch limits on proportional harvest of the estimated biomass for each area. While alternate policy-based allocation formulae are possible, the staff believes that they would have to be consistent with the sustainable yield of the stock and, if the formulae were to have an equitable basis, then they would have to be consistent with the sustainable yield for each regulatory area as well. The staff does not believe such a policy-based approach will be functional unless it has this sustainable basis.

Existing policy-based allocation formulae (e.g. the allocative Catch Sharing Plan (CSP) of the Pacific Fishery Management Council for Area 2A) are implemented after the conservation (sustainability) decision has already been made (i.e., the CSP works entirely within the catch limit adopted external to the CSP). Ultimately, conservation and allocation can be separated but in the hierarchy of decisions, conservation and sustainability must be paramount.

30. How is sublegal bycatch accounted for in CEY calculations?

Total CEY (Constant Exploitation Yield) is calculated by applying a target harvest rate (presently 20%) to estimated exploitable biomass. Fishery CEY in each area (commercial catch in all areas, plus sport catch in Areas 2A and 2B) is calculated by subtracting from total CEY all other removals that are similar to commercial removals in their effect on the stock. These consist of all hook-and-line catches, legal-sized as well as sublegal, and all legal-sized bycatch.

Currently, sublegal bycatch in the trawl fishery is treated differently because it is mainly of fish with a modal length around 50 cm. The effect of these removals on the stock is some years off, and the geographic distribution of the impact is uncertain because migration rates and schedules are unknown. We can calculate that sublegal bycatch at present levels reduces coastwide recruitment at age 8 by 10%, and we include this level of pre-recruit mortality in the fishery simulations that we conduct when evaluating alternative target harvest rates. It turns out that the choice of a target harvest rate is not very sensitive to the level of pre-recruit mortality, but in principle a reduction in sublegal bycatch mortality would increase both recruitment to the stock and total CEY.

Another option is for sublegal bycatch mortality in the halibut setline fisheries to be deducted directly from the CEY for the areas in which it occurs because almost all of it is close to the commercial size limit and, therefore, the impact on total yield to the stock is essentially equivalent to the impact of catching those fish somewhat later in time as legal-sized fish.. To implement such an accounting process would require annual estimates of sublegal size composition in all setline fisheries, as well as mortality rate estimates for the discards by each gear. Comprehensive treatment of this mortality would also require estimation of discards within the recreational fishery, which are substantial in some areas such as Area 3A. Such information does not presently exist and the current process of accounting for such mortality in the setline fishery through harvest rate adjustment is appropriate.

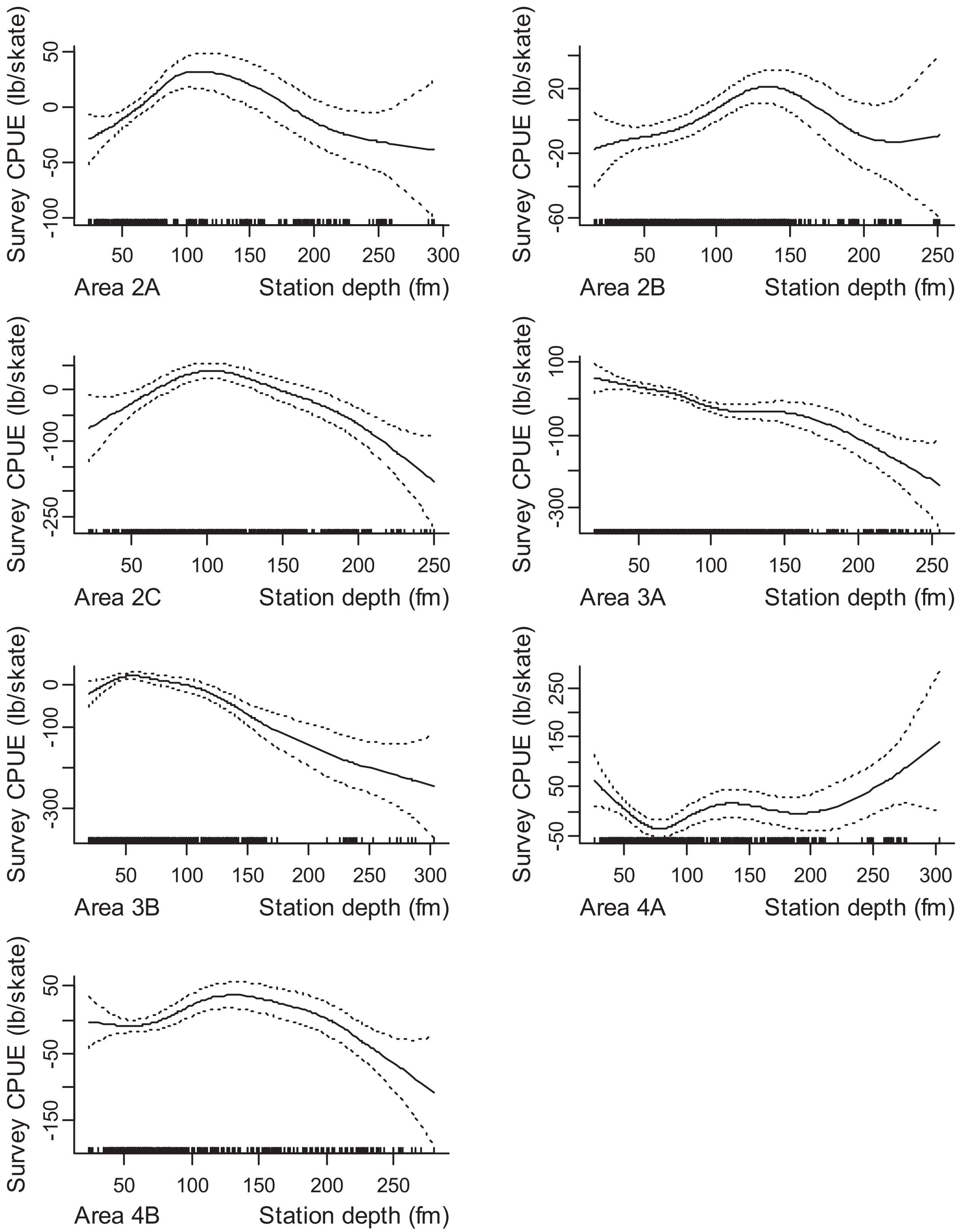

Figure 1. Relationship between depth and survey CPUE, from a GAM model fit (for survey years 2001-2006). The dotted lines show 95% confidence intervals. The ticks on the bottom axis show data points.

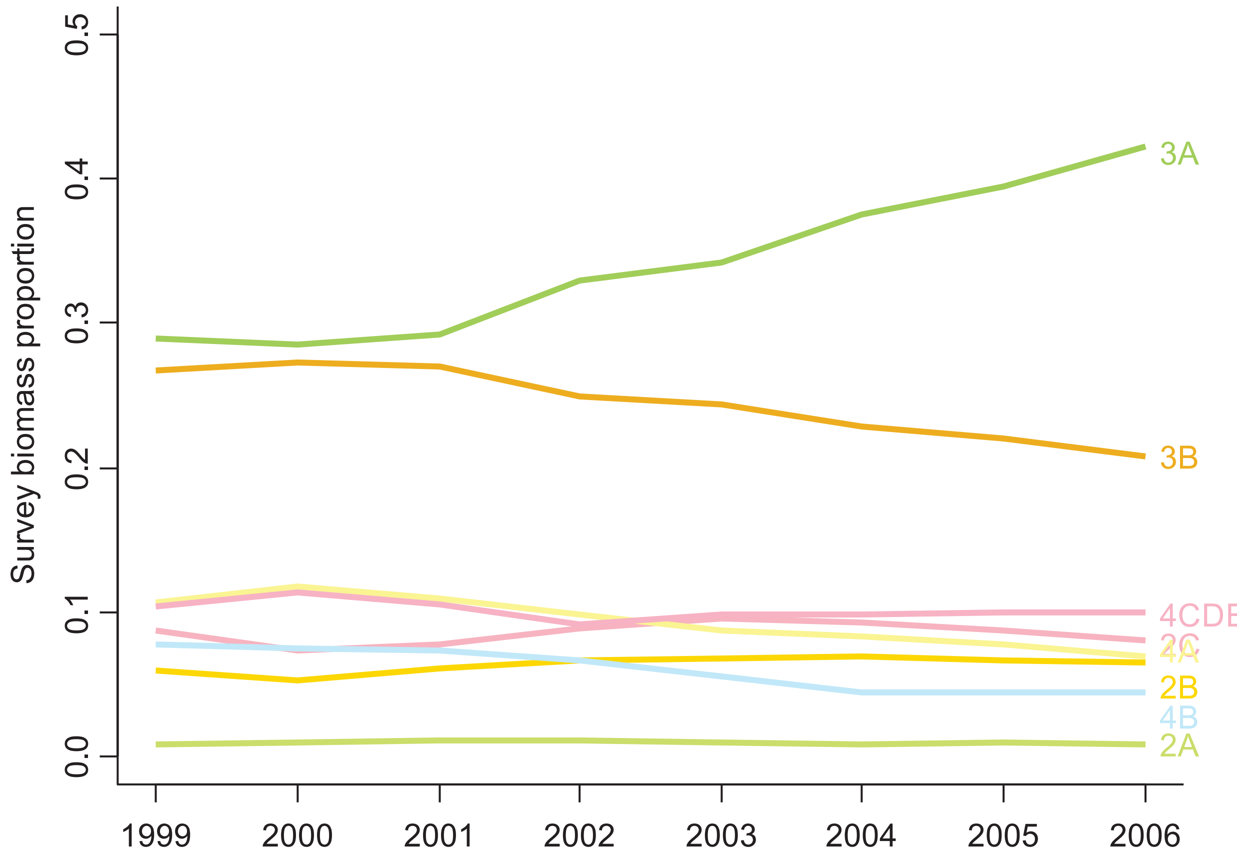

Figure 2. Estimated proportion of exploitable biomass in each area as estimated from bottom area and a 3-year running mean of survey CPUE.

Report on the 2006 Assessment and Harvest Policy of the International Pacific Halibut Commission*

Author

RICC Francis

*This document is presented here in its original format

Prepared for

University of Miami

NIWA Client Report: WLG2007-55
July 2007

NIWA Project: MIA07301

National Institute of Water & Atmospheric Research Ltd
301 Evans Bay Parade, Greta Point, Wellington
Private Bag 14901, Kilbirnie, Wellington, New Zealand
Phone +64-4-386 0300, Fax +64-4-386 0574
www.niwa.co.nz

Contents

Reviewed and Approved for release by:

Dr Rosemary Hurst

Executive Summary

A public workshop was held in Seattle, Washington, June 27-28, 2007 to consider the 2006 stock assessment and harvest policy of the International Pacific Halibut Commission (IPHC). Presentations were made on both the assessment and harvest policy, additional analyses were requested and carried out, and the workshop discussed the results. Following the workshop, the independent reviewers met informally with IPHC staff to discuss technical issues arising from the workshop.

The workshop was well run and IPHC staff were clear in their presentations and helpful in their response to queries.

The following recommendations are offered. High priority should be given to completion of the analysis of the PIT tag data, with the aims of explaining the low recapture rates and improving estimates of migration rates. Future assessment modelling should be spatially structured so that 1) the migration estimates can be used in the assessment, 2) all data can be used in the area apportionment, 3) various current assumptions can be tested, and 4) the interaction between migration and harvest policy can be investigated. There is little evidence to support the survey-based area apportionment procedure, but there is no obvious alternative to it without a spatially structured assessment model. The Commission should consider modifying its harvest policy in response to new information about adult migration. Some suggestions are made which may improve some aspects of the pre-model data analysis, log likelihood, estimation of uncertainty, and any future workshops.

1. Background

This report reviews, at the request of the University of Miami (see Appendix 1), the 2006 assessment and harvest policy of the International Pacific Halibut Commission (IPHC). The author was provided beforehand with links to various documents (Appendix 2) and participated in the workshop which considered the assessment and harvest policy.

2. Review Activities

A public workshop was held at the Nexus Hotel in Seattle, Washington, on June 27-28, 2007 to consider the 2006 stock assessment and harvest policy of the IPHC. The workshop was chaired by Dr Steve Martell (University of British Columbia) and those attending included IPHC staff, several IPHC Commissioners, fishing industry representatives, people from various U.S. and Canadian fisheries agencies and universities, and two external reviewers (including the author).

IPHC staff made a series of presentations to the workshop concerning both the stock assessment and harvest policy. One presentation on the second day reported on additional analyses carried out in response to matters arising on the first day. After the workshop, the two reviewers met informally with IPHC staff to discuss some technical issues arising from the workshop.

3. Findings

Overall, I was impressed by the documents I read, the structure of the workshop and presentations by IPHC staff, and the courtesy and helpfulness of these staff in responding to my queries and those of other workshop participants. They showed high professional standards, as well as strong initiative in detecting and responding to the thorny problem of adult migration.

All my comments and suggestions below should, of course, be considered subject to qualification or rebuttal by IPHC staff. There is a tremendous volume of detail associated with the data and analyses that I was asked to consider, and it was clearly not possible for me to comprehend it all in the time available. I hope that I have correctly identified, and understood, all important details, but I may well have missed some.

The seven parts of this section correspond to the seven questions that I was asked to address (see Appendix 1), which are repeated at the beginning of each subsection. In some places I refer to the documents I was provided by their file names, as given in Appendix 2 (e.g., sa06.pdf, 2k6rara04.pdf).

3.1 Data

Are the stock assessment data adequate? If not, what more is needed?

Five sets of data were used in the assessment model: 1) catch numbers at age and sex; 2) catch per unit effort (CPUE) from the commercial setline fishery; 3) CPUE from the setline survey; 4) proportions at age and sex from the setline survey; and 5) numbers at length in the bycatch.

My impression, after reading the documents and seeing the workshop presentations, is that that these data, and the procedures followed in collecting and processing them, are of the high standard one would expect from an internationally recognised agency like the IPHC. In common with many other agencies, the IPHC found that their earlier age data (from otolith surface readings) were biased and imprecise, but they have dealt with this well using a misclassification matrix.

The most important of the assessment data sets are those from the setline survey. Because they are fishery-independent they are extremely valuable as assessment tools, and I commend the IPHC for its continuing commitment of substantial resources to the survey. I was impressed with the effort IPHC staff devoted, both in design and data analysis, to ensure the quality of the survey data by standardising effort and rejecting stations whose catch rates were affected by predation. During the workshop it was suggested that the grid design of the survey may bias results, and that randomly located stations would be better. I acknowledge this weakness but am, on balance, in favour of the grid design for this survey. Given the large number of stations, any bias is likely to be small, and the grid design reduces variance and provides substantial logistical advantages, and also some analytical advantages which I will discuss in the next section.

3.1.1 Possible data refinements

I have some suggestions concerning refinements in the analysis of the survey data and commercial CPUE. I suspect that these will have little effect on the stock assessment, although a large effect is possible. Even if their effect is small, what is to be gained is a greater confidence that any important patterns in the data are not being overlooked.

With regard to the survey data I have two suggestions. First, the survey area should be stratified for the analysis of both the CPUE indices and proportions at age and sex. I can see nothing to be lost from stratification, and there are two possible gains: removal of any bias that may arise when the number of stations is not exactly proportional to stratum area (slide 11 of Bill Clark's area apportionment talk showed that this is important in Area 2A), and reduction in variance (if there are significant between-stratum differences in catch rates). Depth contours are obvious candidates for stratum boundaries, but I would suggest also using lines perpendicular to the shoreline (e.g., some or all of the lines shown in slide 6 of Claude Dykstra's presentation).

My second suggestion is to carry out an analysis of all the survey data (as suggested by Paul Medley) with the aim of identifying any factors (e.g., time of day, tidal phase, vessel, etc) that might affect catchability. The aims of this analysis are twofold: to increase understanding of setline catchability, and (possibly) improve the survey data used in the stock assessment. For this analysis, the grid design is an advantage because it gives us more power to detect such factors. What I have in mind is something like the following. Define

$$Y_{sy} = \frac{X_{sy}/M_{sy}}{\mathrm{mean}_{y'}\left(X_{sy'}/M_{sy'}\right)}$$

where X_{sy} is the catch rate from station s in year y and M_{sy} is the mean catch rate from all stations in year y that are in the same stratum as station s. Dividing by M_{sy} removes any effects of large-scale spatial and temporal changes in halibut abundance; the other division removes any effect due to the individual station location (the latter is possible only because of the grid design of the surveys). Now use a GLM (or GAM) to look for factors that explain variation in the Y_{sy}, and thus in halibut catchability. The use of standard gear and fishing methods, and the large number of year-station combinations available for this analysis, make this a powerful data set for such an analysis.

For the commercial CPUE data I suggest standardisation using the usual methods, which are described, for example, in several papers in the 2004 review issue of *Fisheries Research* (volume 70, issues 2-3). I am aware that the need for such standardisation in the halibut assessment is limited because the associated catchability is allowed to vary with time. However, it seems to me a bad practice not to remove the effect of any factors that can be shown to affect commercial catchability. It is a difficult task for our assessment models to extract the abundance signal from the considerable noise that is typical in fisheries data, and we should do all we can to help them. Further, such standardisation may improve our knowledge of factors affecting catchability.

Finally, I would like to express my reservations about the use of area weighting to combine the area-specific commercial CPUE indices for use in the coast-wide model. I do not mean to suggest that there is a clearly better way of combining these data. Nor do I want to imply that this method of combination has not been often used in the fisheries literature. I simply wish to point out that area weighting requires two strong assumptions: that the CPUE in the *a*th area, X_a, is proportional to the density of fish in that area (i.e., $X_a = q_a B_a / A_a$, where B_a is the biomass in that area and A_a is its area – in n.m.^{2}, say) (we usually assume X_a is proportional to the *biomass*, rather than the *density*); and that the constant of proportionality, q_a, is the same in all areas. One of my reasons for advocating a different model structure (see Section 3.2.1) is that it both avoids the necessity for these assumptions and allows us to test them.

3.1.2 PIT tag data

The data from the current PIT tag experiment were not used directly in the stock assessment. However, they provided preliminary information on migration which was crucial in the decision to recommend a change in the form of the assessment, from closed-area to coast-wide, with the key result being evidence that a non-trivial proportion of halibut continue to migrate long after recruiting to the fishery. I will argue below that this result also has important implications for the harvest policy (see Section 3.6.3).

I was impressed by the design and scope of this experiment. I think it is comparatively rare that such experiments achieve the ideal of tagging in each area in proportion to abundance as well as the present one. This is particularly important when we see that a high percentage of recaptures from the survey occurred near the tagging location.

There is a pressing need to understand why recapture rates were so low, so that estimated fishing mortalities from the tagging data were only a fraction of those from the assessment. Of particular concern is the possibility that whatever phenomenon is responsible for this under-estimation may bias estimates of migration. However, I do not know how likely this is.

I have a couple of comments on the analysis that attempted to explain the low recapture rates by "small-scale mismatches between the distribution of released tags and the distribution of commercial fishing" (p. 140, 2k6rara04.pdf). First, I do not see how this hypothesis explains low recapture rates. My thinking is that if the tagged fish are restricted to, say, a quarter of the area, then they are available to only a quarter of commercial sets, but within the area they occupy their average density is four times what it would have been had they been distributed across the entire area. Don't the factors of one quarter and four balance out, so that the expected number of tagged fish caught is the same as it would have been had they not been restricted in area? What the mismatch hypothesis *does* imply is that the distribution of numbers of tags per landing should be contiguous, and it is only this which is being tested in the analysis of Table 1 (p. 141, 2k6rara04.pdf). With regard to this analysis, either I have misunderstood it (quite possible!) or the expected numbers in this table are wrong, and the difference between observed and expected numbers is greater than it appears. Here, for the record, are my calculations of the expected numbers for area 2B. I calculated the mean for the Poisson distribution in this area as the total number of tags recovered (242 = 108 + 31x2 + 13x3 + 2x4 + 5x5) divided by the total number of landings (727 = 568 + 108 + 31 + 13 + 2 + 5). Using the Splus function dpois, I calculated the expected numbers in the table as round(727 * dpois(0:5,242/727)), which gives values of 521 173 29 3 0 0. In each area my estimates of the expected number of landings with no tags should be over-estimates, because I treated the observed landings in the column '5+' as containing exactly 5 tags. However, my estimates of these numbers were always less than those in the table.

With regard to the analysis of migration rates, I am concerned that the present analysis, which is on the scale of the regulatory areas, may be biased because it ignores the substantial spatial heterogeneity in scanning rates (see Figure 1, p. 142, 2k6rara04.pdf). I wonder if the analysis could be done on a smaller spatial scale. One approach would be to estimate the following parameters: the probability that an individual will migrate within the next year, and, for those fish that do migrate, the mean and standard deviation of the eastward distance migrated. These parameters could be treated as smooth functions of age and/or length and/or position around the coast. They give a more biologically meaningful description of the migration than does a matrix of between-regulatory area migration rates. That matrix can, of course, be calculated from the above parameters, together with an initial population distribution from the tagging surveys.

There's an important distinction to be made between migration rates, and their effects on the underlying population. For example, preliminary estimates of annual percentage rates of migration between regulatory areas (Table 4, p. 136, 2k6rara04.pdf) suggest approximately equal rates of migration from 3A to 3B (7.7%) and vice versa (9.9%). However, we cannot know the net effect of these migrations without knowing what the pre-migration distribution of the population was. According to both of two alternative estimates of this distribution, the net effect is that Area 3A is a source of fish, whereas 3B is a sink (Table 1). From a management point of view, I think the net effects shown in this table are more important than the migration rates. Table 1 also illustrates the point that the migration may be more complicated than a simple eastward movement (in which Area 3B would not be a sink). I do not mean to suggest here that the migration *is* complicated in this way, because I know that the migration rate estimates on which this table are based are very preliminary. However, it does not do any harm to be reminded that nature is often (usually?) more complicated than we might like to think it is, and halibut migration may not be exclusively eastward.

Table 1: Annual percentage change in population size assuming the migration rates estimated from PIT tag data (see Table 4, p. 136, 2k6rara04.pdf) and two alternative estimates of the pre-migration distribution of the population (from Table 1, sa06.pdf).

Source of pre-migration					Area
population distribution	2B	2C	3A	3B	4A
Closed-area assessment in 2006	13	-5	-5	18	-9
Survey area apportionment in 2006	14	-2	-3	6	-11

3.2 Model structure

Is the structure of the assessment model appropriate? If not, what changes should be made?

Two alternative models were used in the 2006 assessment: closed-area and coast-wide. These models differed only in their input data; their structures were essentially the same. I have no criticisms of this common structure. It dealt well with the available data and allowed the modeller a great deal of flexibility in investigating the shape of selectivity curves and how these curves, and the associated catchabilities, might vary over time.

IPHC staff presented some good arguments against the use of the closed-area model. For example, there is the 2C/3B paradox: "Area 3B is twice the size of area 2C and has a higher survey CPUE, but the [closed-area] assessment says there is more biomass in Area 2C" (p. 83, 2k6rara04.pdf). I am happy to accept these arguments. The problem for me (and for the IPHC, I believe) is that the coast-wide model requires some way of apportioning the estimated current biomass amongst the regulatory areas. It is important to distinguish between accepting the coast-wide model over the closed-area, and accepting the area-apportionment scheme. I know that the IPHC staff are well aware of this distinction but I am not confident that this was true of all workshop participants.

3.2.1 A spatially-structured model

I would like to describe, and advocate, a model that is intermediate between the closed-area and coast-wide models. Its advantages over these models are: 1) it uses information from the PIT tag data; 2) it does the area apportionment in a way that is consistent with both current and historical biomass distributions as well as migration rates; 3) it provides a method of testing various assumptions, including the key one underlying the current area-apportionment scheme (area-independent survey catchability); and 4) it provides a tool to deal with the interaction between migration and harvest policy. Its main disadvantage is that it is markedly more complex, and it might take a year or two before it is stable.

The spatially-structured model divides the population amongst the regulatory areas. The five data sets currently used in the existing models (see Section 3.1) would enter the model separately by area. Thus, for example, there would be a survey and commercial CPUE series for each area. The population dynamics for each regulatory area (i.e., the equations describing how the number of fish in a cohort decline from year to year because of fishing and natural mortality) would be exactly as they are in the existing models with one addition: equations describing movement between areas.

I would expect that migration parameters would be estimated outside the model and then fixed in the model. In this situation, the parameters estimated would be exactly the same as were estimated in the closed-area models (see pp 14-15, sr83.pdf). In fact, we could mimic a simultaneous run of all the closed-area models by setting all migration rates to zero (a useful test of the new model). But this is only the starting point; we can do much better than this with the spatially-structured model. The current closed-area models necessarily estimate different survey and commercial catchabilities and selectivities in each area. With the spatially-structured model we can test a series of hypotheses about between-area differences in these quantities. The obvious hypothesis to test (because it underlies the current area apportionment) is that survey catchability is the same in all areas. If we found (as I would expect) that this hypothesis was rejected when there was no migration, we could ask the question what rates of migration do we need to assume so that this hypothesis is not rejected, and are these migration rates consistent with the PIT tag data? (Bill Clark briefly described – but did not present – informal analyses he did with the current models aimed at addressing precisely this question, treating migration as a change in natural mortality; I believe the spatially-structured model would provide a much more robust, and theoretically sound, approach to this problem).

The spatially-structured model avoids a great weakness of the coast-wide model: its inability to use the considerable quantity of area-specific information. Because management has been by area we might expect that population trends would be different in the different areas. If the current area apportionment is correct these differences would be large, because recent exploitation rates have been very different (see Fig. 6, p. 157, 2k6rara04.pdf). Are these exploitation rates consistent with the area-specific data on catches and catch rates? The coast-wide model provides no way of answering this question, but the spatially-structured model does. Perhaps more importantly, with the spatially-structured model the area apportionment is an integral part of the assessment, and uses all available data. With the coast-wide model, the area apportionment is external to the model, uses only part of the data (the survey CPUE), and, in its current form, requires an untested assumption (area-independent catchability).

The spatially-structured model is also an obvious tool to investigate more fully the effect of migration on the harvest policy, which I discuss in Section 3.6.3.

3.3 Log likelihood

Is the log likelihood used to fit the model appropriate? If not, what should be used?

The objective function lies at the heart of all modern stock assessment models. This function is of great importance because it controls the estimation of parameters (i.e., the model fitting): by definition, the final parameter estimates are those which minimise the value of the objective function. Two types of terms were added together to make the objective function used in the halibut assessment: log likelihoods and penalties (in Bayesian models, a third type – prior distributions – is also used).

The penalty terms used in the halibut assessment seemed to me reasonable and relatively uncontroversial, and so I will say no more about them. Nor will I say anything about the robustification function used to reduce the effect of outliers (Fig. 11, p. 54, sr83.pdf), except that I think it is a good idea.

The role of the log likelihood terms in the objective function is twofold. First, they provide, for each individual observation, a measure of the difference between the observed value, X_{obs}, and the value predicted by the model, X_{model}, given a set of parameter values. The form of this measure is determined by the statistical distribution assumed for each observation. In the halibut model all observations were assumed to be normally distributed, so the measure of difference was always $(X_{\text{model}} - X_{\text{obs}})^2$. The second role of the likelihood terms is to assign a weight to each individual observation, which is a way of telling the model how close X_{model} should be to X_{obs}. A high weight encourages the model to find parameter values which make X_{model} very close to X_{obs}. Most of what I have to say about the halibut log likelihood concerns these weights.

In the halibut assessment these weights were assigned in a three-step process. First, a sampling standard deviation, s, was either calculated for, or assigned to, each observation using a variety of methods (see pp 8-9, sr83.pdf). This assigned a weight of $1/s^2$ to the observation. The model was then fitted using these weights and root mean square errors, τ, were calculated for each type of observation (see Table 1, p. 41, sr83.pdf). The weight assigned to each observation was then changed to $1/(s\tau)^2$, and the model was run again with the new weights. If the resulting fit was considered inadequate for some type of observation, an arbitrary additional factor W was assigned to it to encourage a better fit to it (e.g., a factor of 10 was used for survey and commercial CPUE observations in some assessments – see p. 18, sr83.pdf). Exactly how this factor was applied in the likelihood function was not described, but I believe the effect was to change the weight on these observations to $W/(s\tau)^2$.

Data weighting is often a difficult and controversial issue in stock assessments. It is controversial because different weightings can produce quite different assessment results. It is difficult because it is not possible to provide an objective set of rules that will guarantee the best weighting in all assessments, and so it is hard to avoid a subjective component to the weighting decisions. Subjective weightings lead to the undesirable possibility that the results of assessments carried out by different modellers (with different subjective judgements) could be quite different.

I have several suggestions that I think could improve the halibut assessment by making the data weighting more theoretically (statistically) sound and less subjective. These suggestions may or may not make a significant change to the assessment results, but I think they could make the assessment statistically more sound, and thus more defensible.

3.3.1 Suggested changes to the log likelihood

The approach I suggest is based on viewing the total error associated with each observation, $e_{total} = (X_{model} - X_{obs})$, as being the sum of two parts: $(X_{true} - X_{obs}) + (X_{model} - X_{true})$, where X_{true} is the true value (in the real world, not in the model) of the observed quantity. The first part is what is usually called the observation error, and the second I call process error (some people use the term model error), so we can write $e_{total} = e_{obs} + e_{proc}$. The important point to notice is that although many types of observation contain information about e_{obs} they almost never, by themselves, contain any information about e_{proc}. For example, for survey CPUE we can estimate e_{obs} from the between-station variation in catch rates, but these catch rates contain no information (by themselves) about the random year-to-year variations in survey catchability which I believe form the main contribution to e_{proc}. (I am assuming here that the model assumes no year-to-year variation in this catchability). For the survey proportions at age (and sex), the observations contain abundant information about e_{obs}, but not about e_{proc}, which is often very substantial. One obvious source of process error for this type of observation is the common model assumption that natural mortality does not vary with either year or age. Depending on the type of observation and the assumed sampling distribution, we can describe the size of the two types of errors using either standard deviations (s.d.s, s_{obs} and s_{proc}) or coefficients of variation (c.v.s, c_{obs} and c_{proc}). Whichever approach we use, these quantities add as squares: $s_{\text{total}} = (s_{\text{obs}}^2 + s_{\text{proc}}^2)^{0.5}$ and $c_{\text{total}} = (c_{\text{obs}}^2 + c_{\text{proc}}^2)^{0.5}$. If we apply this view of error structure to the halibut log likelihood it is easy to see that, at least when no additional W factors are used, the halibut s corresponds roughly to my s_{obs}, and τ is associated with s_{proc}.

My first suggestion to improve the log likelihood is to consider changing from the current multiplicative model, $s_{\text{total}} = s_{\text{obs}}\tau$ (or $s_{\text{total}} = s_{\text{obs}}\tau/W^{0.5}$), to a more theoretically sound (I think) additive model – either $s_{\text{total}} = (s_{\text{obs}}^2 + s_{\text{proc}}^2)^{0.5}$ or $c_{\text{total}} = (c_{\text{obs}}^2 + c_{\text{proc}}^2)^{0.5}$. Note that for a given set of observations (e.g., survey proportions at age) there will usually be a different observation error for each individual observation, but a single process error that will be added to all observations within that set. This means that the effect of changing to additive errors will be greatest for sets of observations within which the variation in observation error is large (e.g., proportions, or numbers, at age or length). An example will help to show how different the additive and multiplicative approaches can be. Suppose we base our observation errors for a set of proportions at age on a multinomial distribution with $N = 10\,000$, and the proportions vary from 0.001 to 0.1, so c_{obs} will vary from 0.316 to 0.030, respectively. If c_{proc} is 0.3 (a typical value for my own assessments) then c_{total} will vary from 0.436 to 0.301. Thus, the effect of adding this process error is very substantial for the high proportions (the c.v. increases from 0.030 to 0.301) and much less so for the low proportions (c.v. increase from 0.316 to 0.436). With multiplicative errors and $\tau = 2.5$, the corresponding changes are much greater for the low proportions (from 0.316 to 0.790) and less for the high proportions (from 0.03 to 0.075).

My second suggestion is to check which error model is best. Regardless of any theoretical grounds, the best justification for any approach to modelling error distributions is to show that it is consistent with the data and model. This is fairly easily done with residual plots. For example, with the proportions at age observations, I would standardise all the residuals so that their expected s.d. (according to the assumed error model) is 1, plot the absolute residuals against log proportion, and fit a smooth line through the plotted points. This line should be approximately horizontal if the error model is appropriate. Such an approach can be used to decide between multiplicative and additive errors, and, if additive errors are used, whether to base these on s.d.s or c.v.s.

My third suggestion is to fix the errors for those sets of observations that you wish to be sure of fitting well (primarily the survey CPUE, possibly also the commercial CPUE) and estimate process error within the model for all other sets of observations (i.e., include a c_{proc} or s_{proc} as an estimable parameter for each set – or a τ, if you can show that multiplicative errors are better). I thought that the use of smoother to estimate a total c.v. for halibut commercial CPUE (p.9, sr83.pdf) was an excellent idea, and suggest doing the same for the survey CPUE. I acknowledge there is a subjective element in this, because you have a choice of how smooth to make your smooth line. However, you can at least be upfront about this by including in your assessment document a plot showing the smooth line you fitted to each CPUE data set. That way, it is easy for a reviewer to make a judgement as to whether they believe your line is too smooth or not smooth enough (this is essentially a judgement about how smooth we should expect biomass trajectories to be). Once the model has been fitted you should compare the variance of the CPUE residuals with that predicted by your assumed errors. If this variance is too large it is an indication of a conflict between data sets, or between them and model assumptions.

When faced with this sort of evidence of conflict there are three main responses. The first choice (though often difficult) is to change the model assumptions until the conflict disappears. The second choice is to try to determine which data sets are in conflict and try model runs in which some data sets are omitted. Failing all else, up-weight those observations that you are failing to fit (by reducing s_{total} or c_{total}) and run the model again. The aim is to add sufficient weight to these observations so that the variance of the residuals is about what should be expected according to your original assumed errors (not the up-weighted errors). This up-weighting is not desirable, but at least there is a reasonably objective means of deciding how much additional weight to apply.

My fourth suggestion is to avoid double entering of observations (including total proportions at age, as well as proportions at age for males and females; including CPUE in number as well as in weight; and including CPUE at age, as well as total CPUE). I think I understand why this was done (as an attempt to make sure of good fits to all aspects of the data) but it seems to me quite indefensible from a statistical point of view and it interferes, in a way that I find hard to predict, with the estimation of uncertainty (see Section 3.7).

My final suggestions concern the composition observations (proportions or numbers at age or length). These observations are not derived from simple random samples, so the use of the multinomial distribution to calculate s_{obs} or c_{obs} is inappropriate. For example, the calculation of proportions at age from the setline survey involves two-stage sampling – stations within a stratum, and otoliths within a station – not a simple random sample. A more reliable way of estimating observation c.v.s (or s.d.s) is to use resampling methods on the raw data. Also, I wonder if a normal error distribution is appropriate for the composition data. A normal distribution with a c.v. exceeding about 0.4 will include a reasonable proportion of negative numbers, which do not make sense for composition data. In the composition data I'm used to dealing with c_{total} often exceeds 0.4, so I use a lognormal distribution, rather than a normal. I note that plots like Fig. 10 in sr83.pdf have been used to justify the use of the normal distribution. My guess, but I could be wrong, is that these plots are a mixture of fairly symmetric distributions, for the larger proportions, and positively skewed distributions for the smaller proportions.

3.4 Alternative fits

Is the suite of alternative fits adequate?

I think that the suite of alternative fits presented in the assessment (Table 2, p. 12, sa06.pdf) was perfectly adequate to explore the range of plausible fits to the coast-wide model.

I liked the use of retrospective runs as a diagnostic tool but note that it can be difficult to decide whether a retrospective trend is of concern (indicating a serious fault in the model structure) or simply what could be expected because of a chance pattern of residuals in the observations. Sets of retrospective runs based on simulated data can provide some help in making this decision.

3.5 Area apportionment

Is the area apportionment procedure correct?

For me, this was the weakest part of the assessment.

I do not mean this statement to be a criticism of the IPHC staff, because the approach they took had a logic to it that is hard to escape. They started with the model that had been used previously (the closed-area model) and found that this was inconsistent with the recent information about migration (from the PIT tags). There were also some internal inconsistencies (e.g., the "2C/3B paradox" – see above). Their decision then to switch to a coast-wide model seems utterly reasonable (I doubt that it would have been possible to develop a more complex model, like that in Section 3.2.1, within the time available). Having done that, they needed a method to apportion the total exploitable biomass, as estimated from the coast-wide model, amongst the regulatory areas. The choice they made was the obvious one – to use the fishery-independent survey data – and this required some assumption about the survey catchability in each area. The usual scientific approach is to apply Occam's razor. That is, to make the simplest assumption that is consistent with the data. In this case that assumption is that the survey catchability is the same in all areas. This is not because there was strong (or even much) evidence of equality. Rather it is because there was no plausible way to estimate separate catchabilities by area.

What is the best decision from a scientific point of view is not necessarily the best to use in managing a fishery. I think the Commissioners were prudent to reject the new assessment because it would have led to a substantial change in the allocation of quota amongst regulatory areas. While there were reasonably good grounds to believe that the previous method of allocation (using the closed-area assessment) was flawed, the evidence to support the new method was weak. In such a situation a pragmatic approach is to stay with something like the status quo until the scientific picture becomes clearer.

I have four reasons to doubt the assumption of area-independent survey catchability, though none is strong, and none leads to a clear alternative assumption. The first derives from the area-specific recapture rates for PIT tags in the 2006 survey (bottom part of Table 3, sa06.pdf). A simple chi-square test applied to this table shows a highly significant departure from the assumption of equal catchability (P = 0.005) [I could not understand the assertion that this was only "marginally significant" (bottom of p. 9, sa06.pdf)]. The second reason is the comparison between setline and trawl CPUE in Areas 3A, 3B, and 4A (Fig. 3, prospect.pdf). This was used by IPHC staff to reject the closed-area model, which implied that setline survey catchability was much lower in Area 3A than in 3B and 4A (Fig. 1, prospect.pdf). However, the same data also rejects the assumption of area-independent catchability because, for example, it indicates that setline catchability in Area 3B was 50-60% of that in 4A (the fact that the plotted 3B points in Fig. 3 are lower than those for 4A in all eight length classes makes this difference statistically significant regardless of the associated confidence intervals). I do not consider this strong evidence, because it requires the assumption that trawl catchability is the same in 3A, 3B, and 4A, and I am less willing to make that assumption than the IPHC staff seem to be. The third reason is that the recent Area 2 exploitation rates estimated using the assumption of area-independent survey catchability are very high (see Fig. 6, p. 157, 2k6rara04.pdf), particularly in Area 2B. It seems hard to believe that the effects of such high exploitation rates would not have been noticed before now, particularly by fishers. Finally, I note that catchability is a function of both fisher behaviour and fish behaviour. Much is done to ensure that fisher behaviour is the same throughout the survey area, but it is not possible to standardise fish behaviour. As the environment, biotic and abiotic, changes from Oregon via the Pacific coasts of British Columbia and Alaska to the Bering Sea, I would expect halibut behaviour to change in response, and this change may well affect its catchability.

3.6 Harvest policy

Is the harvest policy appropriate; i.e., does it strike a proper balance between utilization and precaution? If not, how should it modified?

The IPHC management of the halibut fishery in recent years has been based on what is called the Constant Harvest Rate (CHR) policy. The aim of this policy has been to apply the same harvest policy in all areas. In fact, the actual harvest rate has not been constant, either in time or across areas. There are two main reasons for this. First, the target harvest rate has changed over time (varying between 0.20 and 0.35 since 1985) as understanding of the stock dynamics has evolved. This is to be expected, and is not unreasonable. Second, in setting quotas for each regulatory area in each year, various adjustments have been made to the harvest rates. I will comment first on the calculation of target harvest rates, and then on the adjustments to these rates. Finally, I will discuss the effect of migration on this harvest policy.

3.6.1 Calculation of the target harvest rate

I was impressed by the simulation approach used to calculate the target harvest policy (pp 30-36, sr83.pdf). This explicitly allows for periodic regime shifts affecting productivity, density-dependent growth, length-specific selectivity, and time-invariant maturity at age. As a sensible conservative measure, parallel simulations are done assuming that the current low growth rates will persist, even if biomass is reduced. Four performance measures are calculated for each a range of possible target harvest rates: average annual yield, average spawning biomass, average actual harvest rate, and the proportion of years the spawning biomass falls below a threshold (Table 5 and Fig. 29, sr83.pdf). The only potential weakness I could see in these simulations is that they did not allow for various adjustments in harvest rates (i.e., they assumed that the actual harvest rate was the same as the target harvest rate, except for random errors in estimating current biomass).

As is common in simulations of this type, the procedure for choosing the ‘best’ target harvest rate from the results of these simulations involved a trade-off between utilization (as measured by the realised yield) and precaution (as measured by the probability of falling below a biomass threshold). The risk associated with falling below the biomass threshold is an increase in the probability of recruitment failure. The larger the target harvest rate the greater this risk is, but the higher the average yield (assuming no recruitment failure).

I interpret the question at the head of this section as asking whether the IPHC staff found the optimal trade-off between utilization and precaution in choosing the target harvest rate from the simulation results. I can offer no opinion on this because what is optimal depends on how risk-averse the Commission wishes to be, and I have no information about that.

3.6.2 Adjustments to the harvest rate

In the documents I saw I found evidence of three types of adjustment that have, or could have, been made to target harvest rates. The first relates to the reduction that is supposed to occur when the spawning biomass falls below the threshold and/or limit reference points (as illustrated in Fig. 27, p. 73, sr83.pdf). This seems to me a very sensible, and prudent, type of adjustment, though it is a moot point as to whether it is actually part of the IPHC harvest policy (as opposed to the simulation experiment of Section 3.6.1) because, as far as I am aware, it has never been applied.

The second type of adjustment is when IPHC staff either reduce a harvest rate because the assessment is uncertain or there is reason for concern in a particular regulatory region (e.g., the reduction from 0.20 to 0.15 in Areas 4B and 4CDE in 2006), or increase it to reduce the effect of an otherwise large reduction in quota (e.g., the increase from 0.20 to 0.25 in Area 2 in the 2006 coast-wide assessment). While I can understand the reason for making such adjustments I am concerned by their ad hoc nature, particularly when a harvest rate is increased. I could see no reason why the increase in Area 2 should have been to 0.25, and not 0.3, or 0.225.

The third type of adjustment is as a result of the "slow up-fast down" (SUFD) policy (p. 149, 2k6rara04.pdf). I found no precise definition of this policy in the documents I read, but I understood it to mean that the agreed yield for year y, $Y_{\text{agreed},y}$, would be calculated from the recommended yield, $Y_{\text{recommended},y}$, using an algorithm something like the following

$$Y_{\text{agreed},y} = \begin{cases} 0.33Y_{\text{recommended},y} + 0.67Y_{\text{agreed},y-1} & \text{if } Y_{\text{recommended},y} > Y_{\text{agreed},y-1} \\ 0.5Y_{\text{recommended},y} + 0.5Y_{\text{agreed},y-1} & \text{if } Y_{\text{recommended},y} < Y_{\text{agreed},y-1} \end{cases}$$

I was unsure about the status of this policy. I could not understand why it appeared to play no part in the calculation of the yields recommended after the 2006 assessment (Table 1, sa06.pdf). Also, it was unclear to me why, given the existence of this policy, it was deemed necessary to increase the recommended harvest rate in Area 2 from 0.20 to 0.25.

It concerned me that only the first of these three types of adjustment was allowed for in the simulations to calculate target harvest rates. If SUFD is formally part of the IPHC harvest strategy I can see no reason for not including it in these simulations. Perhaps it was not included because it is only intermittently applied? As long as the second type of adjustment is ad hoc, it will not be possible to include it in the simulations. I am not bothered by downward adjustments, because the effect of these is precautionary. However, upward adjustments mean that there is more risk to the stock than is indicated by the simulations which underlie the choice of target harvest rate.

3.6.3 Effect of migration on the harvest policy

If halibut migration rates turn out to be of the order indicated by preliminary estimates from PIT tags the Commission may want to consider modifying their harvest policy. The modification could reduce harvest rates in eastern areas, and increase them in western areas. To see why this change might be advisable we need to consider the rationale behind the current policy.

As I understand it, the harvest policy is intended to protect the spawning stock. Spawning occurs in most, if not all, of the regulatory areas, but it is not known which areas are most important in producing recruits to the coast-wide population. Thus it is prudent to offer the same protection to the spawning population in all areas, and that is achieved by applying the same harvest rate in all areas. One index of the level of protection given is the current spawning biomass expressed as a percentage of what it would have been were there no fishing. To simplify the following discussion I will call this index *PI* (Protection Index). The smaller *PI* is, the less protection is being given to the spawning stock (but note that the index is not linear, so we cannot say that a *PI* of 20 indicates half the protection of a *PI* of 40). The intent of the harvest policy appears to be that, on average, *PI* should be (approximately) the same in all areas.

The effect of migration is that the policy of applying the same harvest rate in all areas no longer achieves the goal of having the same *PI* in all areas. With an eastward migration and equal harvest rates, *PI* will, on average, be lower in the east and higher in the west. This is illustrated in Table 2, which I constructed using results from the migration modelling carried out by IPHC staff (pp 148-149 and Fig. 4, 2k6rara04.pdf) assuming an annual migration rate eastward of 6% (i.e., T = 0.06). For example, fishing with F = 0.2 produces a lower *PI* in Area 2B (24) than in Area 4A (33). The contrast is stronger with a higher rate of fishing: with F = 0.3, *PI* is 16 in Area 2B and 25 in Area 4A. I made two assumptions in calculating these figures: that the total unfished biomass in areas 2B to 4A would be 500 M lb, and that if there were no migration the *PI* for all areas would be 29 at F = 0.2, and 20 at F = 0.3. Neither of these assumptions affect the main point of this table. That is, the ratio of the *PI* values in the two extreme areas (2B and 4A) – 24:33 with F = 0.2, and 16:25 with F = 0.3 – would not change under different assumptions.

Table 2: Equilibrium spawning biomass and protection index (*PI*) by area under three levels of fishing (unfished, *F* = 0.2, *F* = 0.3) assuming an annual migration rate eastward of 6% (*T* = 0.06) and an arbitrary total unfished biomass of 500 M lb. *PI* is the equilibrium spawning biomass expressed as a percentage of what it would be were there no fishing.

	Unfished (F = 0.0)		F = 0.2			F = 0.3		
Area	% of all[1]	M lb	% of all[1]	M lb	*PI*	% of all[1]	M lb	*PI*
2B	23	115	19	28	24	18	18	16
2C	21	105	20	29	28	19	19	18
3A	32	160	35	51	32	36	36	23
3B	16	80	17	25	31	18	18	23
4A	8	40	9	13	33	10	10	25
All	100	500	100	145	29[2]	100	100	20[2]

[1] From Fig. 4, p. 155, 2k6rara04.pdf; [2] Average values from Table 5, p. 44, sr83.pdf

The effect of migration on harvest policy may be greater than is shown in Table 2, because migrations rates may be higher than the 6% I assumed there. In the preliminary estimates of migration rates (Table 4, p. 136, 2k6rara04.pdf) the annual percentage of fish leaving each of areas 4A to 2C ranged from 9% to 15%.

Things are a bit more complicated than shown in Table 2 because the current harvest policy does not actually provide exactly the same protection for the spawning stocks in each area. For example, with a harvest rate of 0.2, *PI* is expected to be 24 in Area 2B, 27 in 2C, and 36 in 3A (Table 5, p. 44, sr83.pdf) (the figure of 29 used in Table 2 is the average of these values). However, the migration modelling on which the calculations of Table 2 was based ignored these differences. I conclude that if we assume no migration, the current policy offers somewhat less protection to the spawning stock in Area 2B than in Area 3B. The effect of migration would be to increase this difference between areas. The question the Commission must address is whether this increase is sufficiently large to be of concern. If so, there will be a need to consider changing the harvest policy.

3.7 Uncertainty

Does the assessment adequately measure and report the uncertainty of the yield recommendations? If not, what more should be done?

The problem of reporting uncertainty in stock assessments and yield recommendations is a difficult one to which there is no simple answer. In discussing this problem I will describe two types of uncertainty (within- and between-model) and then consider whether the correct reporting of uncertainty really matters. To illustrate some points I will use results from some recent assessments of New Zealand hoki.

3.7.1 Within-model uncertainty

The main, and perhaps only, quantitative description of uncertainty presented in the 2006 halibut assessment appears to be a c.v. of 7% for the estimate of current coast-wide exploitable biomass (p. 6, sa06.pdf). It was also reported that this was half the corresponding values for the closed-area assessments. These c.v.s were calculated from the Hessian matrix evaluated at the point estimate. Roughly speaking, the c.v. of 7% implies that we can be reasonably confident that the true exploitable biomass was within +/–14% of the estimated value. Uncertainty in the associated yield recommendations does not seem to have been explicitly quantified. For the closed-area assessments it would be reasonable to assume that the c.v.s for the recommended yields would be similar to those for the exploited biomass. However, for the coast-wide assessment the yield c.v.s would have been much greater than 7% (because of additional uncertainty associated with the area apportionment) but difficult to calculate (because there is no obvious way of quantifying the uncertainty associated with the assumption of area-independent survey catchability).

This use of the Hessian matrix is a standard way of expressing uncertainty in quantities estimated by maximum likelihood. The only criticism I have of the c.v.s presented in the 2006 assessment relate to the changes in the log-likelihood which I suggested in Section 3.3.1. That is to say, I believe we would have more confidence in these c.v.s if these changes were made. I do not support the ad hoc multiplication of 2 to deal with double counting of observations (see bottom of p. 20, sr83.pdf).

C.v.s calculated using the Hessian matrix represent what I call within-model uncertainty. This is hard to define exactly, but roughly speaking it is that uncertainty which exists if the model assumptions are “broadly” true and the specified errors (i.e., s_{total} or c_{total} – see Section 3.3.1) are correct. The looseness of this definition derives from the word "broadly", which is needed because the process errors (represented by s_{process} or c_{process}) make some allowance for departures from the model assumptions.

Unfortunately, different approaches to estimation can produce very different estimates of within-model uncertainty. For example, some stock assessments use Bayesian estimation, rather than maxim likelihood. A joint posterior distribution is estimated for all model parameters and the uncertainty in any estimated quantity (such as exploitable biomass) is calculated from this posterior. In one Bayesian assessment I did for New Zealand hoki I compared Hessian-derived variances for all model parameters with those from the Bayesian posterior and found that these sometimes differed by more than an order of magnitude in either direction, though the Bayesian variances were usually larger.

3.7.2 Between-model uncertainty

In some stock assessments it is clear that not all the uncertainty can be expressed by estimates of within-model uncertainty from a single model. A good example of this is provided by the 2006 assessment of two stocks of New Zealand hoki, in which three alternative models produced quite different views of the range of uncertainty in certain biomass estimates (Figure 1). The main differences between the three models concerned the method of explaining the relative lack of old fish (either domed selectivity or age-dependent natural mortality) and whether there is natal stock fidelity (i.e., whether a fish always belongs to the same stock as its parents). In this assessment none of the three models was clearly superior to the others in either plausibility or fit to the observations.

Some measure of between-model uncertainty was presented informally in the 2006 coast-wide halibut assessment in the table of estimates from seven alternative models (p. 5, sa06.pdf), although some of these models were clearly markedly inferior to others, and so should be excluded for this purpose.

It is not easy to be confident in any particular assessment that the range of models considered captures all significant between-model uncertainty. Given sufficient time, it is very likely that the IPHC staff could have constructed other, more complex models which extended this uncertainty. One obvious example is the spatially-structured model I advocate in Section 3.2.1 (though I repeat that I do not think that this could have been easily done within the time frame of the 2006 assessment).

There are some drawbacks to presenting between-model uncertainty to fisheries decision makers. These decision makers often require a single estimate of each quantity of interest (e.g., current exploitable biomass), and a single c.v. representing the total uncertainty in that quantity. Unfortunately, it is usually not possible to produce such single estimates from multiple models. To do this requires assigning a weight to each alternative model, which is far from straightforward.

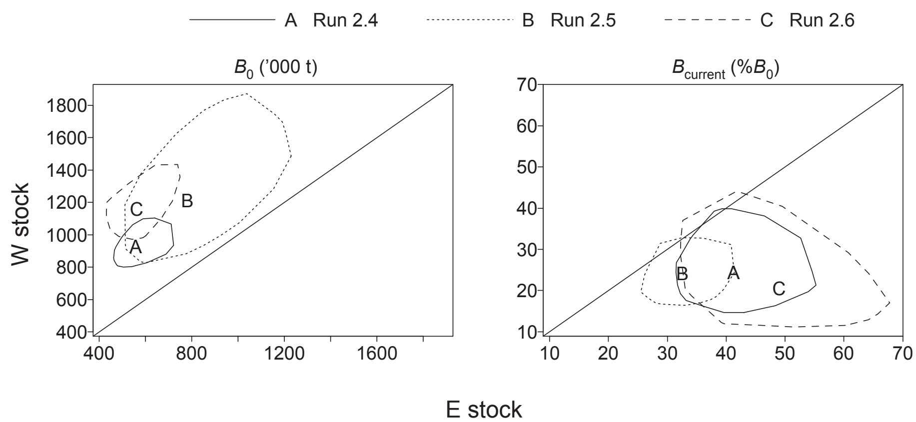

Figure 1: Illustration of between-model uncertainty in the 2006 assessment of two stocks of New Zealand hoki. Each panel shows point estimates ('A', 'B', 'C') of either unfished biomass (B_0, left panel) or current biomass ($B_{current}$ as % B_0, right panel) for the two stocks (E and W) and the associated approximate 95% confidence regions (polygons) from three alternative models: 2.4, 2.5, and 2.6.

3.7.3 Does uncertainty matter?

In most fisheries agencies charged with producing stock assessments it goes without saying that it is desirable, where possible, to provide some measure of uncertainty for key estimates, such as current biomass and yields. I think it also widely acknowledged that our ability to estimate uncertainty is usually poor. This view was well expressed by IPHC staff: "For lack of anything better, such estimates [i.e., Hessian-based estimates of variance] are often reported anyway" (p. 20, sr83.pdf). I have tried in the preceding discussion to provide some explanation for this situation.

However, my view is that our inability to estimate uncertainty well probably does not matter very much. Consider, for example, what might have happened if the PIT tagging had not been done, so that there was no reason to doubt the closed-area assessment in 2006. The Commissioners would have been provided with yield estimates for each regulatory area, and these would have had c.v.s of about 15%. The question is, is it likely that the quota decisions made by the Commissioners would have been different had these c.v.s been different (say 5%, or 25%)? I suspect not.

4. Conclusions

4.1 The seven questions

Are the stock assessment data adequate? If not, what more is needed?

The data currently used in the assessment model seem adequate, although I have some suggestions for refinements in how the survey data and commercial CPUE are analysed outside the model (Section 3.1.1). There is a pressing need to complete the analysis of the PIT tag data, with the aims of explaining the low recaptures rates and improving the estimation of migration rates (Section 3.1.2).

Is the structure of the assessment model appropriate? If not, what changes should be made?

The structure of the assessment model (essentially the same in the closed-area and coast-wide assessments) was appropriate given the time frame of the 2006 assessment.

For future assessments I suggest the development of a spatially-structured model (described in Section 3.2.1) which has the following advantages: 1) it uses migration information from the PIT tag data; 2) it uses all data (rather than just survey CPUE) to do the area apportionment; 3) it provides a method of testing various assumptions, including the key one underlying the current area-apportionment scheme (area-independent survey catchability); and 4) it provides a tool to deal with the interaction between migration and harvest policy.

Is the log likelihood used to fit the model appropriate? If not, what should be used?

I offer several suggestions to improve the log likelihood (Section 3.3.1). These suggestions may or may not make a significant change to the assessment results, but I think they could make the assessment statistically more sound, and thus more defensible.

Is the suite of alternative fits adequate?

Yes.

Is the area apportionment procedure correct?

While there is clear evidence that the previous area apportionment procedure (using the closed-area assessments) is flawed, there is little evidence that the new procedure is correct, and some grounds to believe that it is not. Unfortunately, there was no obvious alternative at the time of the assessment. The spatially-structured model offers an alternative for the future.

Is the harvest policy appropriate; i.e., does it strike a proper balance between utilization and precaution? If not, how should it modified?

The simulation procedure used to support the selection of a target harvest rate is sound. I cannot say whether the selected target harvest rate strikes “a proper balance between utilization and precaution” because what is “proper” depends on what level of risk is acceptable to the Commission, and I have no information on that. I found the practice of adjusting harvest rates for each year and area (Section 3.6.2) hard to understand, and note that upward adjustments compromise the target harvest rates derived from the above simulation procedure.

Does the assessment adequately measure and report the uncertainty of the yield recommendations? If not, what more should be done?

My suggestions for improving the log likelihood should also improve the current estimate of stock assessment uncertainty, but I note that this type of uncertainty is not well estimated anywhere.

4.2 Suggestions for any future workshops

I have two small suggestions that may be of use in planning any future workshops. First, it makes the reviewers’ task much simpler when material presented in workshop is (at least mostly) restricted to that in documents provided beforehand (the obvious exception being analyses requested during the workshop). Second, I thought that 1.5 days was the absolute minimum duration, given the amount of material to present, digest, and discuss. Longer would have been better.

APPENDIX 1: Statement of Work

This appendix contains the Statement of Work that formed part of the consulting agreement between the University of Miami and the author.

General

The International Pacific Halibut Commission seeks an independent review of its stock assessment and harvest policy. The assessment is an age- and sex-structured model, coded in AD Model Builder, which is similar in most respects to the groundfish assessments done by the NMFS Alaska Fisheries Science Center. The harvest policy is based on stock and fishery simulations that include environment-dependent recruitment and density-dependent growth as reported in previous published analysis. The reviewers should be fully competent in modern stock assessment methods, in particular the use of AD Model Builder software and contemporary statistical catch-at-age analysis.

Specific

The reviewer's work will be as follows:

1. The reviewer will read Scientific Report 83, which describes the stock assessment and harvest policy in detail.

http://www.iphc.washington.edu/halcom/research/sa/papers/sr83.pdf

2. The reviewer will read the 2006 stock assessment documents.

http://www.iphc.washington.edu/halcom/research/sa/papers/sa06.pdf
http://www.iphc.washington.edu/halcom/research/sa/papers/prospect.pdf
http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara04.pdf

3. The reviewer will attend a public assessment workshop in Seattle, Washington, from June 27-28, 2007, where the material in Scientific Report 83 will be presented and discussed in detail with attendees from other agencies, the industry, and the public. The IPHC will arrange an independent chair for this workshop.

4. The reviewer will meet with IPHC staff during the meeting to go over any questions arising during the meeting.

5. No later than July 13, 2007, the reviewer will submit an independent report via electronic mail to Dr. David Die (ddie@rsmas.miami.edu) and Mr. Manoj Shivlani (mshivlani@rsmas.miami.edu). The report must include, but not be restricted to, answers to the following specific questions:

(i) Are the stock assessment data adequate? If not, what more is needed?
(ii) Is the structure of the assessment model appropriate? If not, what changes should be made?
(iii) Is the log likelihood used to fit the model appropriate? If not, what should be used?
(iv) Is the suite of alternative fits adequate?
(v) Is the area apportionment procedure correct?
(vi) Is the harvest policy appropriate; i.e., does it strike a proper balance between utilization and precaution? If not, how should it modified?
(vii) Does the assessment adequately measure and report the uncertainty of the yield recommendations? If not, what more should be done?

APPENDIX 2: Materials Provided

Before the workshop the reviewer was provided with the following links to relevant documents.

Necessary Material

The workshop is intended be technical, rather than educational, and participants are expected to be familiar with the IPHC assessment process. Relevant documents are IPHC Scientific Report 83, which describes the assessment model and harvest policy:

http://www.iphc.washington.edu/halcom/research/sa/papers/sr83.pdf

as well as documents describing the 2006 IPHC stock assessment and yield recommendations:

http://www.iphc.washington.edu/halcom/research/sa/papers/sa06.pdf

http://www.iphc.washington.edu/halcom/research/sa/papers/prospect.pdf

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara04.pdf

http://www.iphc.washington.edu/halcom/research/sa/papers/indepth.pdf (new addition)

http://www.iphc.washington.edu/halcom/research/sa/papers/hook.pdf (new addition)

Background Material

General reference list on the Commission

Please note the Commission's website contains all historical publications in our Scientific and Technical report series, as well as most Reports of Assessment and Research Activities from 1991-2006. Literature on the IPHC website is at:

http://www.iphc.washington.edu/halcom/literatu.htm

Pertinent general references on Commission mandate, history, and management.

McCaughran, D.A. and S.H. Hoag. 1992. The 1979 Protocol to the Convention and Related Legislation. 32 p. http://www.iphc.washington.edu/halcom/pubs/techrep/tech0026.pdf

The Pacific Halibut: Biology, Fishery, and Management. 1998. http://www.iphc.washington.edu/halcom/pubs/techrep/tech0040.pdf

The 2007 Pacific Halibut Fishery Regulations.

http://www.iphc.washington.edu/halcom/pubs/regs/2007iphcregs.pdf

Kong, T.M., H.L. Gilroy, and R.C. Leikly. 2004. Definition of IPHC statistical areas. 72 p.

http://www.iphc.washington.edu/halcom/pubs/techrep/tech0049.pdf

Clark, W.G., B. A. Vienneau, C. L. Blood, and J. E. Forsberg. 2000. A review of IPHC catch sampling for age and size composition from 1935 through 1999, including estimates for the years 1963-1990. 40 p.

http://www.iphc.washington.edu/halcom/pubs/techrep/tech0042.pdf

IPHC Grid Surveys

Soderland, E, C.L. Dykstra, T. Geernaert, A.M. Ranta, and E Anderson. 2007. 2006 Standardized stock assessment survey Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006: 335 - 365.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara09.pdf

References on Commercial Fishery

Commercial catch

Forsberg, J.E. 2001. Aging manual for Pacific halibut: procedures and methods used at the International Pacific Halibut Commission (IPHC). Int. Pac. Halibut Comm. Tech. Rep. No. 46: 56 p.

Page down to Technical Report 46

http://www.iphc.washington.edu/halcom/pubs/technical.htm

Forsberg, J.E. 2005. Age distribution of the commercial halibut catch for 2006. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006: 75–80.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara03.pdf

Gilroy, H.L., J.E. Forsberg and W.G. Clark. 1995. Changes in commercial catch sampling and age determination procedures for Pacific halibut, 1982 to 1993. Int. Pac. Halibut Comm. Tech. Rep. No. 32: 44 p.

http://www.iphc.washington.edu/halcom/pubs/techrep/tech0032.pdf

Gilroy, H.L., L.M. Hutton, and K.A. Gravel. 2007. 2006 commercial fishery and regulations changes. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006: 35 – 46.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara03.pdf

Hutton, L.M. and K.A. Gravel. 2007. Commercial catch sampling. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006: 67- 73.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara03.pdf

Wastage from the commercial halibut fishery

Gilroy, H.L. 2007. Wastage in the 2006 Pacific halibut fishery. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006: 55 -58.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara03.pdf

References on Other Removals

Bycatch

Clark, W. G., and S. R. Hare. 1998. Accounting for bycatch in management of the Pacific halibut fishery. No. Amer. J. Fish. Mgmt. 18:809-821.

http://www.iphc.washington.edu/Staff/hare/html/papers/bycatch/abst_byc.html

Williams, G. H., C. C. Schmitt, S. H. Hoag, and J. D. Berger. 1989. Incidental catch and mortality of Pacific halibut, 1962-1989. Int. Pac. Halibut Comm., Tech. Rep. No. 23. 94 p.

http://www.iphc.washington.edu/halcom/pubs/techrep/tech0023.pdf

Williams, G. H. 2007. Incidental catch and mortality of Pacific halibut, 1962-2006. Report of Assessment and Research Activities 2006:163-174.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara05.pdf

Sport fishery

Blood, C. L. 2007. 2006 sport fishery. Report of Assessment and Research Activities 2006:47-54.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara03.pdf

Meyer, S. C. 2003. Recreational halibut fishery statistics for Southcentral Alaska (Regulatory Area 3A), 1995-1999. A report to the International Pacific Halibut Commission. AK Dept. Fish and Game, Spec. Pub. No. 03-06, Anchorage.

http://www.sf.adfg.state.ak.us/FedAidPDFs/Sp03-06.pdf

Personal Use (Subsistence)

Fall, J. A., Koster, D., and Davis, B. 2006. Subsistence harvests of Pacific halibut in Alaska, 2005. AK Dept. Fish and Game, Tech. Paper 320. 182 p.

http://www.subsistence.adfg.state.ak.us/TechPap/tp320.pdf

Williams, G. H. 2007. The personal use harvest of Pacific halibut in 2005. Report of Assessment and Research Activities 2006:59-62.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara03.pdf

References on PIT Tag Recoveries

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

http://www.iphc.washington.edu/halcom/pubs/rara/2005rara/2k5rara04.pdf

Clark, W.G. 2006b. Possible causes of low PIT tag recovery rates in 2004. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2005:285-292.

http://www.iphc.washington.edu/halcom/pubs/rara/2005rara/2k5rara07.pdf

Clark, W.G. 2007. Further investigations of low PIT tag recovery rates. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:139-144.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara04.pdf

Forsberg, J.E. 2007. Portside sampling for recovered PIT tags in Pacific halibut. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:277-297.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara08.pdf

Webster, R. A. and Clark, W.G. 2007. Analysis of PIT tag recoveries through 2006. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:129-137.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara04.pdf

References on Harvest Policy

Clark, W. G., and S. R. Hare. 2007. Motivation and plan for a coastwide stock assessment. Report of Assessment and Research Activities 2006: 145-160.

http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara04.pdf

Clark, W. G., and S. R. Hare. 2006. Assessment and management of Pacific halibut: data, methods and policy. Scientific Report 83: 45-65

http://www.iphc.washington.edu/halcom/pubs/scirep/SciReport0083.pdf

Clark, W. G. 2004. Effects of gear type, hook spacing and hook size on commercial selectivity and catchability. Report of Assessment and Research Activities 2005: 145-150 on Area 4B.

http://www.iphc.washington.edu/halcom/pubs/rara/2005rara/2k5rara04.pdf

Clark, W. G., and S. R. Hare. 2004. Updated simulation analysis of the CCC harvest policy with separate accounting of males and females. Report of Assessment and Research Activities 2004: 171-183 on YPR and 185-197 on Area 4CDE.

http://www.iphc.washington.edu/halcom/pubs/rara/2004rara/2k4RARA05.pdf

Links to more in-depth material

http://www.iphc.washington.edu/halcom/pubs/outside/Clark_Hare_1998.pdf (Treatment of bycatch)

http://www.iphc.washington.edu/halcom/pubs/outside/Clark_Hare_2002.pdf (Climate impacts on halibut growth and recruitment)

http://www.iphc.washington.edu/halcom/pubs/outside/Clark_Hare_2004.pdf (The CCC harvest policy)

http://www.iphc.washington.edu/halcom/pubs/outside/mantua_hare_2002.pdf (The PDO)

UM Independent System for Peer Reviews
Consultant Report on: International Pacific Halibut Commission (IPHC) stock assessment and harvest policy review*

Paul A. Medley[1]

June 27-28, 2007

*This document is reproduced here in its original format

[1]Paul Medley
Sunny View
Jack Hole
Alne, YO61 1RT
U.K.
paul.medley@virgin.net

Contents

Executive Summary

The data are adequate for the stock assessment applied. However, the CPUE indices could be improved through standardisation using GLM or GAM models. A similar analysis of the survey indices could be used to correct for various factors not accounted for by the survey design, and lead to a better understanding of how CPUE might relate to abundance.

The model fits the data very well, and there is probably little that can be done to improve the fit. The log likelihood used to fit the model is appropriate, although each data should only be represented once in the likelihood if possible. The uncertainty is generally reported well. The main problem is that the current assessment has significant retrospective bias, which the assessment should seek to reduce.

It is not clear what is causing the retrospective bias, but it is likely at least partly to be due to spatial dynamics. Given the tagging results, it makes sense to have moved from the closed-area assessments to a coastwide assessment. However, this remains an approximation as the tagging movement rates are not high enough to suggest rapid mixing and the fishing fleets do not move freely across the region. Therefore a further improvement would be to explicitly model the areas within a coastwide model. At the very least, this would allow examination of a significant source of uncertainty. Separating the gear catchability from local density of each population component could allow population mapping[1], so that in the longer term, a spatial model reflecting population structure as well as administrative areas might be developed.

Are the stock assessment data adequate?

> The stock assessment data are adequate, but the abundance indices would be improved through standardisation, removing effects on catch rates not related to abundance, such as depth, substrate type and tide (see Stock Assessment Data and Abundance Indices).

Is the structure of the assessment model appropriate?

> The structure of the assessment model is appropriate for a coastwide assessment, and the coastwide assessment appears to be a more accurate description of the population than previous area specific assessments. However, a further improvement would be to use a model of separate population connected through migration estimating from tagging returns (see Model Structure and Likelihood, Tagging Experiment, Modelling Local Density).

Is the log likelihood used to fit the model appropriate?

[1] This refers to how the fish population is distributed in terms of the density of its components, such as age, size, sex and maturity, to geographical areas.

The log-likelihood used to fit the model is appropriate for the aggregated data used. Each observation should only be included in the likelihood once, if possible (see Model Structure and Likelihood).

Is the suite of alternative fits adequate?

The alternative fits presented in the report and at the meeting were adequate in assessing the sensitivity of the model to various formulations with the exception of the area specific versus coastwide assessment (see Model Structure and Likelihood, Tagging Experiment).

Is the area apportionment procedure correct?

Given the coastwide assessment, the area apportionment procedure is reasonable, but would be improved if the biomass in each area could be estimated within the assessment (see Harvest Policy and Area Apportionment Procedure).

Is the harvest policy appropriate?

The harvest strategy strikes the proper balance between utilization and precaution (see Harvest Policy and Area Apportionment Procedure). The current policy has rebuilt and maintained the stock at a healthy level.

Does the assessment adequately measure and report the uncertainty of the yield recommendations?

The assessment adequately measures and reports the uncertainty in yield (see Harvest Policy and Area Apportionment Procedure).

Introduction

An independent review of the International Pacific Halibut Commission's (IPHC) stock assessment and harvest policy was carried out. The assessment is an age- and sex-structured model, coded in AD Model Builder. The harvest policy is based on stock and fishery simulations that include environment-dependent recruitment and density-dependent growth as reported in previous published analysis.

The documents for review were made available through the IPHC web site (http://www.iphc.washington.edu). A meeting of the reviewers, assessment scientists and public was arranged for 27-28 July 2007. IPHC scientists presented the stock assessment at the meeting. An opportunity to ask questions was afforded to the reviewers on 29 July 2007. The following independent review report covers the documents and issues arising during the meeting and discussions.

Summary of Findings and Recommendations

Stock Assessment Data and Abundance Indices

The data are adequate for the stock assessment applied. The catches appear to be complete and the main discards are accounted for. The model makes good use of the extensive survey and commercial CPUE and the "break-and-burn" method produces reliable age estimates. The survey is particularly valuable as it systematically covers the population and provides a platform for other initiatives, such as the tagging programme. However, the data treatment outlined below could improve accuracy in the assessment.

The CPUE indices could be improved through standardisation. Various factors which could bias the survey index are not accounted for. Although the effect of standardising for these factors may be small in the survey index, developing a standardisation method should prove to be important for the commercial CPUE.

The survey data may be biased dependent on various factors which the survey cannot control for, notably substrate type, tides, moon state, and temperature. It is logistically not possible to apply true random sampling in time of survey sites, so it is wise to make some correction for these factors, which could also lead to a better understanding of the behaviour and distribution of the population.

Standardisation will require breaking down the data into individual sets, so zero catches may occur in some strata. Dealing with set-by-set data suggests an alternative likelihood to the normal, bounded by, but allowing for zero. While a robust binomial-type likelihood while suffice with two possibilities, multiple hook states (such as, halibut catch, bait present, bait lost, or other species) require a multinomial-type likelihood.

It would make sense, at least in the first instance to use a generalised linear model to standardise the survey data. The simplest approach would be to fit a linear predictor containing catch type interaction terms for substrate, tide, temperature and so on, to the longline set catches using a Poisson quasi-likelihood, where the likelihood is conditional on the number of hooks set (see McCullagh and Nelder 1989; pp. 209-213). The catch types other than halibut are nuisance parameters, but may affect the perceived abundance. Although the analysis conducted by the scientists suggests competition for hooks is unlikely to make much difference, any difference it does make can be corrected for during the standardisation process.

Model Structure and Likelihood

The model fits the data very well, and there is probably little that can be done to improve the fit. However, a good fit to data does not automatically mean that the assessment models all the underlying processes well. The survey and commercial data contain limited or no information on many population processes, such as natural mortality and movement rates.

The main problem for the current assessment is the significant retrospective bias. Checking the ability of an assessment to forecast outcomes is its most important test. Retrospective bias may be due to problems with the model structure, but also could be due to the past data time series. With longer time series, it is possible the bias will be reduced. It is also possible that the bias can be reduced by modelling population processes which are not adequately described in the current model. Currently the assessment relies on the age data to identify recruitment, and by itself this does not appear to be adequate. Assessing whether changes have improved the model should be based on whether they reduce the retrospective bias.

The most likely problem is poor representation of the spatial dynamics. Adding space to population models can introduce much more complexity, so the simplest representation of population spatial structure is desirable. Significant improvements in the current assessment are likely only if spatial structure of the population is more explicitly taken into account. The tagging returns suggest significant rates of movement, but not so overwhelming that the population can be considered fully mixed even after a few years. There is a clear migration of juveniles from west to east for spawning, and the selectivity appears to be different for the same gear in different areas. These changes in selectivity are most likely due to different densities for the different age and sex components of the population.

The assessment has changed from separate assessments for each administrative area to a single coastwide assessment. The coastwide assessment assumes that there is a single mixed population. Based on the life history information and on movement rates inferred from the tagging, it is apparent the separate stocks assumption does not hold, and the coastwide assessment is likely to be a better description of the population, although it remains an approximation to the real spatial dynamics.

A better model might allow explicit modelling of areas connected through immigration. The simplest formulation is to have a fixed emigration rate between areas, so that the rates are directly comparable with the mortality rates. This leads to a set of simultaneous linear differential equations, one for each area.

$$\frac{dN_i}{dt} = \sum_j \alpha_{ji} N_j - N_i \left(\sum_j \alpha_{ij} + M + F_i \right) \quad i \neq j$$

Other migration models could be used, but this form allows a smooth transition between the closed-area and coastwide assessments. The solution to the equations[2] is exactly analogous to the standard negative exponential population model:

$$\mathbf{N_{t+1}} = \mathbf{Exp(-Z)\ N_t}$$

$$\mathbf{Z} = \begin{bmatrix} -M - F_1 - \alpha_{12} - \alpha_{13} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & -M - F_2 - \alpha_{21} - \alpha_{23} & \alpha_{32} \\ \alpha_{13} & \alpha_{23} & -M - F_3 - \alpha_{31} - \alpha_{32} \end{bmatrix}$$

$$\mathbf{C_t} = \mathbf{F\ Z^{-1}}\left(\mathbf{N_t} - \mathbf{N_{t+1}}\right)$$

Where $\mathbf{F}$ = vector of fishing mortalities among areas, $\mathbf{N_t}$ vector of population size at time t, $\mathbf{M}$=vector natural mortalities and α_{ij} is the emigration rate from area i to j.

The only complication is the exponent of the matrix, which can be found using the Taylor expansion (there are better computationally more efficient methods). This requires a considerable number of matrix multiplications, which might be considered acceptable for a small number of areas. For larger number of areas where $\mathbf{F}$ is changed frequently an approximation could be used, where the catches are removed half way through the time period:

$$\mathbf{N_{t+0.5}} = \mathbf{Exp}(-\mathbf{A}/2)\ \mathbf{N_t}$$

$$\mathbf{C_t} = \left(\mathbf{1} - \mathbf{Exp}(\mathbf{F})\right)\mathbf{N_{t+0.5}}$$

$$\mathbf{N_{t+1}} = \mathbf{Exp}(-\mathbf{A}/2)\ \left(\mathbf{N_{t+0.5}} - \mathbf{C_t}\right)$$

Where in this instance, the $\mathbf{F}$ values form the trace of the matrix, so the e^{-F} can be evaluated by taking the exponent of the individual values. Therefore the emigration / natural mortality matrix $\mathbf{A}$ need only be calculated once. This might also avoid the need for age specific matrices, assuming migration is not age dependent.

Where the movement parameters are equal to zero (α_{ij}=0), the assessment should produce the same results as the previous area specific assessments. However, by combining the data, the assessment should produce better estimates of parameters, particularly those common to all areas. Conversely, very high movement parameters (α_{ij}>>F) should produce an assessment very similar to the coastwide assessment currently proposed. In either case, selectivity and recruitment can vary by area.

[2] The solution can be found in any text book dealing with simultaneous linear differential equations with constant coefficients.

This type of model is more complicated to fit than either the area-specific or coastwide models, but allows much greater flexibility in the way the areas are treated. It also should make the apportionment procedure straightforward as the assessment will automatically give estimates of the biomass by area.

This spatial model is simple enough that it can be used for the assessment model. However, in the first instance, this or a similar model might be used to evaluate potential problems in the coastwide assessment by simulating data where movement rates are commensurate with the tagging data.

The alternative fits presented in the report and at the meeting were adequate in assessing the sensitivity of the model to various formulations with the exception of the area population structure described above. A further test which may be worth considering is a length-dependent natural mortality (Lorenzen, 2005), particularly as the sexes exhibit different growth. It is not likely that this change have a large impact unless the population being modelled includes the smallest animals, but this alternative may be a more realistic representation of how natural mortality changes with size.

Tagging Experiment

The difference between the tagging and the model estimates of exploitation rate are a significant source of uncertainty. This is a potentially important issue as the difference could provide a clue on how to improve the assessment. The theories considered so far, primarily looking at sampling bias, have been found not to apply.

The most obvious concern is that tag-induced mortality is significant and/or tag loss occurs. Various experiments of double tagging and holding tagged animals suggest these factors will not be significant.

Another option is that the overall biomass has been underestimated by the model. It is possible to imagine a “cryptic biomass”, which has lower catchability than the fully exploited population, and tagged animals move between these exploited populations. The stock assessment model, which relies on measurable depletion, may estimate a smaller stock than actually there and overestimate exploitation rates. This hypothesis would favour the tagging estimates over the assessment’s estimates of exploitation rates.

It is most likely that the tagged animals were taken predominantly from the more catchable population. This would lead to higher estimates of mortality for the tagged animals, whereas the opposite is observed. Based on this differential change in catchability, it is difficult to propose a realistic hypothesis which could explain the difference between the observed and expected returns.

Therefore, perhaps the most likely explanation is that tagged fish become “hook shy”. Hook shy fish, having been hooked once, are less likely to be hooked again, and is, in essence, an artefact of the tagging experiment. If this is true, it suggests that the stock assessment model gives better estimates of the exploitation rates. It may be worth considering alternative methods to catch halibuts, such as traps, for future tagging experiments. Otherwise the current treatment of the uncertainty associated with the difference between the experiment and model is adequate.

Modelling Local Density

It may be possible to model local population on a finer scale than administrative areas, using the survey data and generalized additive models to estimate density. This would involve better understanding of the detailed interaction between the fishery and the local population, but lead to improved spatial representation of the stock. Small scale experiments, such as depletion experiments would provide data describing the underlying processes. A successful depletion experiment should provide the local abundance estimates for the population by sex and age. The experiment population needs to be closed, or the migration needs to be estimated. It would also be wise to have a fishery independent abundance index, such as video transects from remote observation vehicles.

More detailed data may allow development of alternative models of density and catchability, which are based on theories on how the population behaves rather than statistical descriptions of data. For example, on the small scale, catchability changes with distance between hooks, suggesting that the halibut population is distributed more evenly than random. If fish are relatively static, the chance that a fish is hooked will depend largely on the area over which a hook will be attracting fish. The area that a hook effectively fishes needs to be adjusted by the amount the areas overlap.

$$A = 2\pi r^2 - 2r^2 \operatorname{Arcsin}\left(b / r\right) - 2ab$$

Where r = radius of the effective area fished, a = 0.5 times the distance between hooks and $b = (r^2 - a^2)^{0.5}$. Clearly, when $r<a$ there is no intersection and the area becomes equal to the full circle area with radius = r. This gives rise to a pattern close to that observed from variable hook distance experiments (Fig. 1). The asymptote for catchability is reached when hooks are about 40ft apart, so the effective fishing area radius is about 20ft (6.0m), and each hook fishes an average 113m^2. Independence between hooks is desirable in the survey, therefore separating hooks so that they do not interfere is better as it increases the survey efficiency. The estimate should allow a local density estimate of the stock size to be calculated, which may become more important as more sophisticated spatial models are developed.

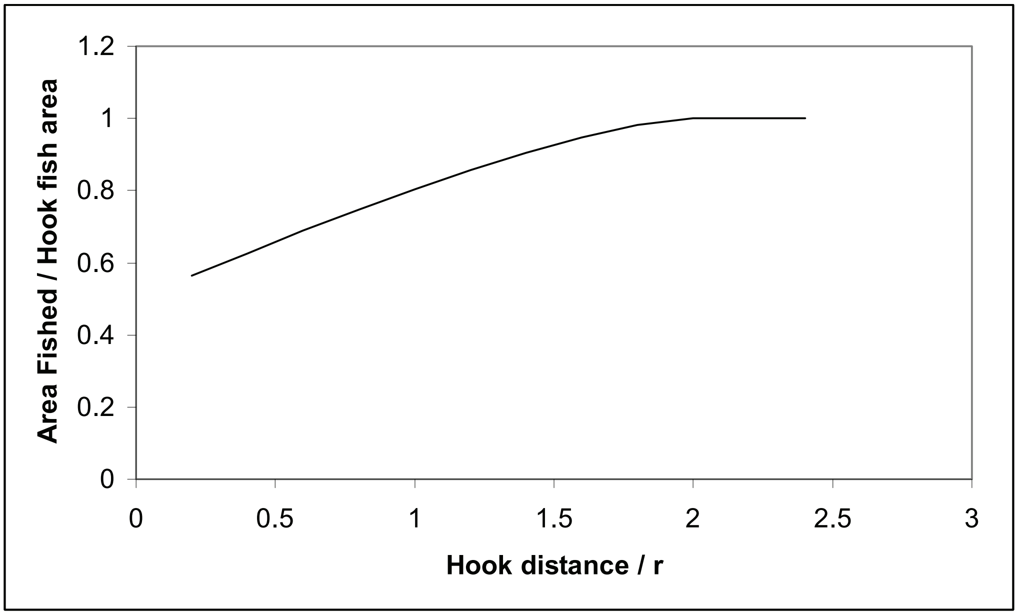

Figure 1 The asymptote is reached when the hook distance is twice the radius of the effective area fished. If this pattern is observed, it strongly suggests that catchable fish are distributed evenly and move relatively slowly.

Detailed modelling of gear catchability would allow the separation of gear selectivity from the site specific densities. The survey could be used to generate a map of any consistent patterns in population structure across the region and develop a spatial population model which would reflect the population structure as well as the fishery administrative areas.

Harvest Policy and Area Apportionment Procedure

The harvest strategy is to apply approximately the same level of exploitation in the different areas across the population range. A total allowable catch is recommended based on the overall estimated biomass, and distributed among areas based on the survey, corrected for catchability. It was suggested that an equal exploitation rate in all areas would be precautionary. This is reasonable until critical life history habitats and life stages can be identified. However, given adequate testing, exploitation rates could be allowed to vary from area to area. For reasons of conservation, exploitation rates could be reduced in the main spawning areas, and for economic reasons, higher exploitation rates could be allowed in areas with highest fishing capacity.

The current area apportionment procedure is a reasonable way to allocate quota, given the coastwide assessment. The procedure is sensitive to survey catchability in particular, and allocation not based on raw survey catch rates may lead to decisions which test the credulity of the fishermen.

The current procedure has the important attribute that the exploitation rate will decrease in an area as its relative biomass decreases, but poor estimates of q may lead to instability. Underestimating q will overestimate the local biomass and contribute to depletion. With a coastwide stock assessment, such local depletion will not be automatically detected.

Overall the harvest policy can be justified and is precautionary, and the assessment adequately measures and reports the uncertainty in yield. However, the current and alternative procedures should be tested through a management strategy evaluation which would allow the harvest strategy to be more clearly tested against various model assumptions.

References

Lorenzen, K. (2005) Population dynamics and potential of fisheries stock enhancement: practical theory for assessment and policy analysis. Philosophical Transactions of the Royal Society of London. Fisheries Theme Issue 2004

McCullagh P, Nelder JA (1989) Generalized linear models. Second Edition. Chapman and Hall, New York.

Appendix I: Document List

The following the main documents used in the review and were available from the IPHC web site (http://www.iphc.washington.edu).

William G. Clark and Steven R. Hare Assessment and management of Pacific halibut: data, methods, and policy.

William G. Clark and Steven R. Hare Assessment of the Pacific halibut stock at the end of 2006.

William G. Clark and Steven R. Hare Motivation and plan for a coastwide stock assessment.

William G. Clark Further investigations of low PIT tag recovery rates.

William G. Clark Effects of gear type, hook spacing, and hook size on commercial selectivity and catchability.

William G. Clark Effect of hook competition on survey CPUE.

William G. Clark Effect of station depth distribution on survey CPUE.

William G. Clark Possible causes of low PIT tag recovery rates in 2004.

Appendix II: Statement of Work

International Pacific Halibut Commission (IPHC) stock assessment and harvest policy review

June 14, 2007

General

The International Pacific Halibut Commission seeks an independent review of its stock assessment and harvest policy. The assessment is an age- and sex-structured model, coded in AD Model Builder, which is similar in most respects to the groundfish assessments done by the NMFS Alaska Fisheries Science Center. The harvest policy is based on stock and fishery simulations that include environment-dependent recruitment and density-dependent growth as reported in previous published analysis. The reviewers should be fully competent in modern stock assessment methods, in particular the use of AD Model Builder software and contemporary statistical catch-at-age analysis.

Specific

The reviewer's work will be as follows:

1. The reviewer will read Scientific Report 83, which describes the stock assessment and harvest policy in detail.

http://www.iphc.washington.edu/halcom/research/sa/papers/sr83.pdf

2. The reviewer will read the 2006 stock assessment documents.

http://www.iphc.washington.edu/halcom/research/sa/papers/sa06.pdf
http://www.iphc.washington.edu/halcom/research/sa/papers/prospect.pdf
http://www.iphc.washington.edu/halcom/pubs/rara/2006rara/2k6rara04.pdf

3. The reviewer will attend a public assessment workshop in Seattle, Washington, from June 27-28, 2007, where the material in Scientific Report 83 will be presented and discussed in detail with attendees from other agencies, the industry, and the public. The IPHC will arrange an independent chair for this workshop.

4. The reviewer will meet with IPHC staff during the meeting to go over any questions arising during the meeting.

5. No later than July 13, 2007, the reviewer will submit an independent report via electronic mail to Dr. David Die (ddie@rsmas.miami.edu) and Mr. Manoj Shivlani (mshivlani@rsmas.miami.edu). The report must include, but not be restricted to, answers to the following specific questions:

(i) Are the stock assessment data adequate? If not, what more is needed?
(ii) Is the structure of the assessment model appropriate? If not, what changes should be made?
(iii) Is the log likelihood used to fit the model appropriate? If not, what should be used?
(iv) Is the suite of alternative fits adequate?
(v) Is the area apportionment procedure correct?
(vi) Is the harvest policy appropriate; i.e., does it strike a proper balance between utilization and precaution? If not, how should it modified?
(vii) Does the assessment adequately measure and report the uncertainty of the yield recommendations? If not, what more should be done?

Staff response to the CIE reviewers' reports

William G. Clark, Steven R. Hare, and Ray A. Webster

Abstract

An external review of the annual IPHC stock assessment was conducted in June 2007. This paper summarizes the reviewers' findings and recommendations, and sets out the staff's response to them.

Background

The IPHC arranges periodic external reviews of the annual stock assessment. One was done in 1997 after a major change in assessment methods in 1995, and another was done in 2007, after further changes made in the 2003 and 2006 assessments. For the 2007 review, the Commission contracted with the Center for Independent Experts (CIE) at the University of Miami, which is contracted to supply external reviewers for NMFS assessment reviews. The Center recruited Drs. Chris Francis and Paul Medley to review the halibut stock assessment. They were given various background papers to read, attended a two-day public workshop in Seattle in June (IPHC Staff 2008), and met with staff the following day to discuss remaining questions. This paper is the staff's response to the reviewers' comments and recommendations.

Overview of reviewers' findings

The terms of reference called for the reviewers to answer seven questions about the assessment and harvest policy. This section summarizes their answers and recommendations. Subsequent sections discuss their comments and recommendations for further work in detail. Our seven questions and their answers were as follows:

1. Are the stock assessment data adequate? If not, what more is needed?

Both reviewers were impressed by the quantity and quality of the data that go into the assessment, and we believe rightly so. The Commission puts a lot of effort into collecting good commercial and survey data, and these data are the real strength of the assessment. Both reviewers approved of the grid design of the survey, believing the potential for bias noted during the workshop was minimal. They recommended that the survey and commercial CPUE data be standardized, as discussed below. Chris Francis also recommended estimating the sampling variances of age compositions by resampling rather than by applying the simple random sampling formula.

The reviewers were concerned, as we are, by the lack of agreement in most regulatory areas between the stock assessment and the PIT tag data (which are not used in the assessment). They stressed the importance of resolving the conflict and made some recommendations concerning the PIT tag analysis.

2. Is the structure of the assessment model appropriate? If not, what changes should be made?

The reviewers were content with the structure of the model used to do both the closed-area and coastwide assessments, and they agreed that the change to a coastwide assessment was appropriate in view of the likely bias in the closed-area assessments due to ongoing migration of recruited fish. They both recommended that the staff develop a spatially structured coastwide model that would include migration among areas and would be fitted to area-specific data, much like the closed-area assessments.

3. Is the log likelihood used to fit the model appropriate? If not, what should be used?

Chris Francis suggested that we try out an additive rather than the present multiplicative model of process error, and consider setting some of the error terms to fixed values, as detailed below. Paul Medley observed that in view of the very good fits the log likelihood was not likely to be a serious issue, but the poor retrospective behavior of the fits needed attention. The staff agrees.

Both reviewers were critical of the double fitting of some data types, e.g. fitting the model to the commercial CPUE at age/sex as well as the commercial catch at age/sex.

4. Is the suite of alternative fits adequate?

The reviewers were content with the range of alternatives shown at the workshop.

5. Is the area apportionment procedure correct?

Paul Medley found the area apportionment of the coastwide biomass estimate based on survey data to be “reasonable”. Chris Francis called it “the weakest part of the assessment” because he saw little evidence to support the assumption of equal survey catchability among areas and some reasons to doubt it. He did not suggest an alternative procedure in the short term but believed the spatially-structured coastwide model could eventually provide more reliable estimates of biomass in each area given good estimates of migration rates.

6. Is the harvest policy appropriate; i.e., does it strike a proper balance between utilization and precaution? If not, how should it be modified?

The reviewers were mostly satisfied with the simulations used to choose a target constant harvest rate, particularly with the inclusion of environmental and density dependent effects on productivity. Chris Francis pointed out that if the “slow up, fast down” (SUFD) adjustment is really standard procedure, it should be built into the simulations. He also questioned the ad hoc precautionary reductions of the target harvest rate in some areas, and raised the issue of whether harvest policy should be engineered so as to produce the same proportional reduction in spawning biomass in all areas. (In the presence of migration, a constant fishing mortality rate in all areas will reduce spawning biomass in the destination areas more than in the source areas.)

Paul Medley thought that the harvest policy struck a proper balance; Chris Francis did not regard the question as a technical issue, but rather a policy issue for the Commissioners.

7. Does the assessment adequately measure and report the uncertainty of the yield recommendations? If not, what more should be done?

Paul Medley thought the assessment adequate in this respect. Chris Francis commented that there is no reliable way to estimate the uncertainty associated with the choice of one model rather than another, which is typically much larger than the uncertainty associated with the fit of a given model. But he stated his view that “our inability to estimate uncertainty well probably does not matter very much” because managers aren’t greatly influenced by variance estimates anyway.

Discussion of specific comments and recommendations

Analysis of PIT tag data

Chris Francis had some questions about an analysis (Clark 2007) that sought to check for small-scale mismatches between the distributions of tag releases and commercial catches. He makes the sensible argument (bottom of p. 4 of his report) that even if the tagged fish are restricted to a fraction of the total area (in the neighborhood of survey stations), their density in those places should be correspondingly higher than it would be if they were distributed throughout the area, and the numerous tag recoveries in commercial catches in the occupied area should therefore exactly compensate for the lack of recoveries elsewhere. This would be true if commercial catches were taken from occupied areas in proportion to the size of those areas, but there is no assurance that they are. Commercial fishing effort is not uniformly or randomly distributed. It takes place at certain customary locations that may or may not overlap the neighborhood of survey stations in proportion to their size. Unfortunately our commercial logbook data do not allow for a direct examination of the degree of overlap on a small scale, but we cannot rule out small-scale mismatches.

In the same paragraph, he reports an apparent error in the calculation of the expected number of landings with 0, 1, 2, 3, 4, and 5 or more tag recoveries that was done to check for contagion in the distribution of tag recoveries. He appears to have misunderstood the published numbers, which were obtained by calculating the null Poisson distribution of tag recoveries for each landing and then summing the probabilities over landings to obtain the expected totals. This is the correct procedure to follow because it allows for the variation in numbers of fish scanned among landings. His alternative calculations are based on a single Poisson distribution for all landings.

On a more important point, Chris was concerned that "the present analysis, which is on the scale of regulatory areas, may be biased because it ignores the substantial heterogeneity in scanning rates", and suggested doing the analysis on a smaller spatial scale. It is certainly true that unaccounted for heterogeneity in detection can be a source of bias in capture-recapture estimation, and allowing for variation in scanning rates among the smaller IPHC statistical areas is therefore a worthwhile suggestion in principle. In practice, however, estimating the scanning rates for statistical areas is not straightforward, as the statistical area of the scanned fish is rarely recorded. However, for most landings, we are able to match the scanning data with independently recorded landed weight data, which is divided into fish harvested per statistical area. Therefore, to estimate how many fish were scanned in each statistical area for each landing, we must make the assumption that scanning was done in proportion to the fraction of the landing from each statistical area. Likewise, in assigning a statistical area to a recovered tag, the best we can do is choose the area that contributed the greatest proportion of the landing's weight. Another complicating factor is that not all scanned landings can be matched to landed weight data, possibly due to minor discrepancies in either the scan or landing data files. This means we must also adjust for those unmatched landings in estimating scanning rates. Our concern is that in assigning statistical areas to recovered tags and in estimating scanning rates by statistical area, we may be introducing another source of bias into the estimation, particularly for areas with relatively few landings.

Nevertheless, it is an interesting modeling exercise, and we are working on tag-recovery models with migration among statistical areas in a manner similar to that suggested by the reviewer. For Gulf of Alaska statistical areas only, from 4A to 2B we have 34 statistical areas from west to east, and the models assume a truncated discrete normal distribution for the distance (number of statistical areas; can be negative) a fish moves in a single year. Coding of the models is currently in progress.

Standardization of survey and commercial CPUE

Both reviewers recommended that we standardize survey and commercial CPUE by fitting GLM/GAM models to estimate and adjust for the effects of covariates such as depth, tide stage, and distance from shore. Chris Francis offered detailed suggestions on a way of rescaling the survey CPUE to remove station effects and trends.

The staff does in fact have an analysis of survey CPUE in the works that will accomplish the purpose, but without rescaling the data. Rather than attempting to remove spatial and temporal effects prior to analysis and then assume the observations are independent, we plan to account for spatial and temporal dependence directly in the modeling. This will allow us to model and map the changing distribution of halibut CPUE over time, as well as model covariate effects on catch rates in a manner that correctly accounts for the study design.

As noted by the reviewers, standardization of commercial CPUE will not affect the assessment because commercial catchability is allowed to drift over time when the model is fitted. At present we partially standardize commercial CPUE by adjusting for hook spacing and gear type (fixed or snap). We could do more, and it would be interesting for its own sake, but we do not regard this analysis as necessary for the assessment because of the allowed drift in commercial catchability.

Estimation of the sampling variance of age compositions

At present we treat the survey and commercial age samples as simple random samples from the catch, but as Chris Francis noted they are not. They are subsamples from clusters of the catch taken by particular vessels at particular times and places, so the estimates have a larger variance than given the random sampling formula. He suggests estimating the variances by resampling the data.

Our commercial and survey age samples are obtained by sampling large numbers of commercial landings and survey catches at very low rates so as distribute the sample evenly over the catches. They have some of the character of systematic samples, so in fact they may be less variable than random samples. In any case Chris' point is well taken, and we will at least do some trial variance estimates using resampling methods to see how they compare with the simple random sampling variances.

Development of a spatially-structured coastwide model

In view of the likely bias in the closed-area assessments due to migration, both reviewers agreed with the use of a combined coastwide dataset to estimate coastwide abundance in the 2006 assessment. They pointed out, however, that if migration rates were known, it would be possible to build a spatially structured coastwide model that included migration among regulatory areas, and to fit this model to area-specific data, just as in the closed-area assessments. They recommended that this sort of assessment be developed.

The staff agrees that this sort of model is feasible. In fact a migratory coastwide model fitted to area-specific data was one of the three standard assessment methods used in the 1980s. The problem with this sort of model is how to estimate the migration rates, which must be estimated externally to the model. In the 1980s migration rates were estimated from recoveries of external tags, which are not reliable for that purpose. The reviewers (and the staff) hope that the present

PIT tag experiment will eventually provide useful estimates of migration rates, but at present the PIT tag experiment is presenting us with more questions than answers. The reviewers suggest that even without reliable estimates of actual migration rates, it would be instructive to build and fit a spatially structured model with a range of migration rates to see what it produced in the way of area-specific estimates of abundance, catchability, and selectivity.

There is no question that the area-specific abundance estimates obtained with this sort of model would be largely determined by the migration rates supplied to it. We have investigated the size of these effects by doing the closed-area assessments with higher and lower rates of natural mortality, because the effect of a given rate of net migration on the number of fish in a given area is exactly the same a change in natural mortality by that amount. What we found, and reported at the 2007 annual meeting, is that even low rates of net migration result in large changes in abundance estimates. For example, running the Area 3B closed-area assessment with an assumed net emigration rate of only 5% increases the present biomass estimate by nearly 60%. Running the Area 3A assessment with an assumed net immigration rate of 5% reduces the present biomass estimate by 25%. In retrospect, we can see that we should have included a full presentation of these results at the workshop.

Even taking an optimistic view of the migration rate estimates that might eventually come from the PIT tag analysis, we do not believe that they will be sufficiently reliable to enable us to make reliable estimates of area-specific biomass by fitting a spatially-structured, migratory coastwide model. We expect there will always be some questions about the reliability of the data, and further questions about the variation of migration rates with size, year-class strength, and perhaps sex. We therefore do not believe that building and fitting such a model would ever be useful for recommending catch limits. Nor do we see that it would provide much more information about the effect of migration on area-specific estimates than we have obtained by doing the closed-area fits with a range of fixed natural mortality rates.

Additive vs multiplicative modeling of process error

The log likelihood that is calculated to fit the present model uses Fournier's (1990) system of variance scalers. That system was adopted because it is published, it is widely accepted, and in our case it does in fact produce distributions of residuals that are quite close to standard normal distributions. It is nevertheless somewhat puzzling in that it makes process error a multiple of sampling error for each data point, whereas one might sensibly think that sampling and process errors should be additive rather than multiplicative.

Chris Francis recommended that we try estimating an additive process error variance for each data type rather than a variance scaler, and outlined some diagnostics for deciding between the additive and multiplicative models. This is a good idea, and we will pursue it.

Fixing some error variances

Chris also suggested that we fix the error variance of key data types—the ones we want to be sure of fitting well—and estimate the rest. For example, we could run a data smoother through the survey CPUE series, calculate the variance around the smoother, and fix that when fitting the model.

This procedure would have the effect of giving more weight to the selected data series, but in a way that is less subjective than simply applying an arbitrary weight when calculating the sum of squares. We are not averse to doing that in order to achieve a satisfactory fit to the CPUE data series, and in fact did so in the closed-area assessments. In the coastwide assessment, neutral weighting of all data types results in good fits to all data series. Changing the weights, as was done in the custom model fits reported at the workshop, had almost no effect on the biomass estimates. Thus while we would certainly fix some error variances or use unequal weights at need, we do not see the need to do so in the coastwide assessment at present.

Double fitting of some data types

In the production assessment some data are fitted twice: the model is fitted to commercial catch at age and to commercial CPUE at age, and it is fitted to the total catch at age as well as to the female and male catch at age. Both reviewers criticized this practice as redundant, and in the custom model fits done at the workshop all double fitting was eliminated with no effect on the biomass estimates.

While conceding that double fitting is redundant, the staff believes that it may be doing some good and isn't doing any harm, at least not to the point estimates. (It obviously invalidates the Hessian-based variance estimates, as we recognized in our published and public reports of the assessment.) If we had deterministic data, then fitting the model to just the catch at age/sex and the total CPUE would give exactly the same estimates as fitting it to the catch at age/sex and the CPUE at age/sex and the total CPUE. But we do not have deterministic data. We have stochastic data, so the two fits are not the same. Which is better? We want the model to fit the catch at age/sex well because the estimates of year-class size depend on that. But we also want the model fit CPUE at age/sex well to make sure it is tracking cohorts. Why not fit to both?

In short, we acknowledge the statistical drawbacks of double fitting as regards variance estimation, but we think it may be a good idea when locating point estimates. This is an issue that needs some research. It is not of any practical importance in the halibut assessment, as demonstrated at the workshop.

Survey apportionment of the coastwide biomass estimate

Chris Francis listed "four reasons to doubt the assumption of area-independent survey catchability, though none is strong, and none leads to a clear alternative assumption". We too have doubts about survey catchability being equal in all areas, and we expect those doubts to persist unless and until we find some independent and indisputable way of measuring halibut density. But we think Chris overstates the evidence against the assumption and understates the evidence in favor, so without wanting to be overly argumentative we would like to comment on his four reasons.

1. The 2006 survey recoveries show significantly different recovery rates among areas. True, but that was a total of only 60 fish. The 1200 or so commercial recoveries in the same table showed no difference in recovery rates among Areas 2B, 3A, and 3B, and a higher rate in Area 2C.

2. The comparative trawl/setline data show a higher setline survey catchability in Area 4A than in Area 3B. The figure in question is reproduced as Figure 1 below. Obviously the data are noisy. However, they clearly show that the ratio of setline to trawl CPUE in Area 3B is neither 2-3 times what it was in Area 3A nor is it consistently higher in Area 3B, which were the important points in considering whether the closed-area assessments were credible. Similarly, Area 4A shows higher ratios for smaller fish but the same ratios as Area 3A for larger fish. We intend to do further analysis of the trawl/setline survey data, but we do not believe that this figure supports a firm conclusion that setline survey catchability is lower in Area 3B than in Area 4A.

3. If exploitation rates were really as high as the assessment estimates in Area 2B, there would be a noticeable depletion during the course of the fishery. Since the workshop we have examined seasonal variation in commercial CPUE in each regulatory area, and we do see a depletion effect in Area 2B.

4. You can standardize the survey but not the fish, and there are probably differences in fish behavior over the large range of halibut. We agree, and for that reason worry most about catchability at the extremes of the range. And the analysis of hook competition data from the survey did indicate some differences in Area 2A and in Areas 4B and 4D. But across the bulk of the range, from Area 2B through Area 4A, there was no important difference in the intensity of hook competition.

Incorporation of the "slow up/fast down" procedure into the harvest rate analysis

The slow up/fast down (SUFD) procedure was introduced by the staff in 1999 as a way to formalize and standardize what had been the traditional Commission practice of phasing in change in catch limits, although it was never presented as a standalone IPHC policy. The SUFD procedure was used selectively between 2000 and 2004 to limit radical changes in catch limits when the estimates of available yield differed substantially from the catch limits of the previous year. Between 2005 and 2007, the SUFD procedure was applied in all areas except Areas 4B and 4CDE (where extra precaution was believed warranted).

It is a trivial exercise to code an SUFD policy into the harvest rate simulations. Chris expressed support for the downward adjustment aspect of the policy, i.e., not taking the full increase in yield when a catch limit was going to increase. He was less keen on the upward adjustments, i.e., not taking the entire decrease when a catch limit was going to decrease. This would argue for a SU policy, also a trivial matter to code and test. Presentation of the harvest policy is already a complicated matter with layers of tables and alternate states of nature to test and analyze. However, since the SUFD procedure has been consistently applied in the staff's catch limit recommendations, an analysis will be prepared and presented for the next Annual Meeting.

Adjusting for the effect of migration when setting area-specific harvest rates

Chris Francis makes note of the fact that application of equal harvest rates in all areas has the effect of disproportionately lowering spawning biomass – relative to the unfished distribution - in the eastern areas. The effect is magnified at higher migration and higher fishing mortality rates. An imbalance in applied harvest rates, as the staff believes to have been the case, with higher rates in the east and lower rates in the western areas, further reduces relative spawning biomass in the eastern areas. Chris notes that one means of balancing spawning biomass reduction equally in all areas would be to apply imbalanced harvest rates with higher rates in the western areas than in the eastern areas. The amount of imbalance would be affected by the assumed migration rate – both the annual rate as well as which ages continue an eastward migration.

Changing the harvest policy such that higher harvest rates are actually applied to the western areas than in the eastern areas – subject to an overall coastwide harvest rate – would require an even greater change from the current situation than was proposed by staff at the last Annual Meeting. Chris' reconstructed area Protection Index (PI, essentially the reduction rate in area specific spawning biomass at various harvest rates) assumed a migration rate of 6%. In the simulations which he references (2007 RARA, pp. 148-149), this 6% migration rate was an annual rate (moving one regulatory area east) for all fish ages 8-20. The 6% value is likely on the high side of migration rates and it is not yet clear that fish continuously migrate up to age 20. At lower migration rates and assuming a lower fishing mortality rate (F=0.2), the difference in PI's among areas is much smaller than those illustrated by Chris, **if equal harvest rates are applied in each area.** The unbalanced harvest rates seen the past 10-20 years have the impact stated by Chris of lowering the PI disproportionately in the eastern areas.

Computation of the exact imbalance of harvest rates to apply by area is subject to the same difficulties faced by a stock assessment that attempts to incorporate migration rates. We see this as a particularly perilous course, given our knowledge of those migration rates The simulations will be sensitive to the rates and those rates are currently poorly estimated. It does appear that the eastern areas are capable of still being productive despite being fished at a higher rate than the western areas. However, part of that sustained production has resulted from the historical lower harvesting rate in the western areas. As harvest rates have increased to the west there is less biomass available to "fill in" the biomass removed in the east. Implementing a policy of equal harvest rates in all areas (save those areas of special concern) has the effect of mostly maintaining a distribution of spawning biomass similar to the unfished state while not grossly disturbing the recent distribution of harvests, at least as compared to the situation of reversing the imbalance of harvest rates and implementing higher rates to the west.

Other comments on the IPHC harvest policy

Chris notes that a potential weakness of the simulations is that it was assumed the actual harvest rate was always the same as the target harvest rate, except for random errors in estimating current biomass. One set of simulations that is done assumes error in current biomass, however, the error is autocorrelated thus making it less random. Thus, within the simulations, errors in estimating biomass translate into errors in applied harvest rate and the autocorrelation is intended to mimic the process whereby stock assessment estimation errors are correlated through time.

A note is made to the ad hoc increase in recommended harvest in Area 2 to 0.25 for the 2007 catch limit recommendations. This is neither part of the simulations nor part of the SUFD procedure. This one-time, ad hoc procedure was implemented as a means of transitioning to the substantially lower catch limits in Area 2 that resulted from the combined effect of the changes in assessment and apportionment. Staff reasoning was that the area had sustained high harvest rates for many years and was not on the verge of collapse and the intent was to move towards the desired harvest rates without having to go the entire way in one year. This line of reasoning is comparable to that behind the SUFD procedure.

References

Clark, W. G. 2007. Further investigations of low PIT tag recovery rates. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:139-144.

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

IPHC Staff. 2008. International Pacific Halibut Commission Stock Assessment Workshop. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2007:99-120.

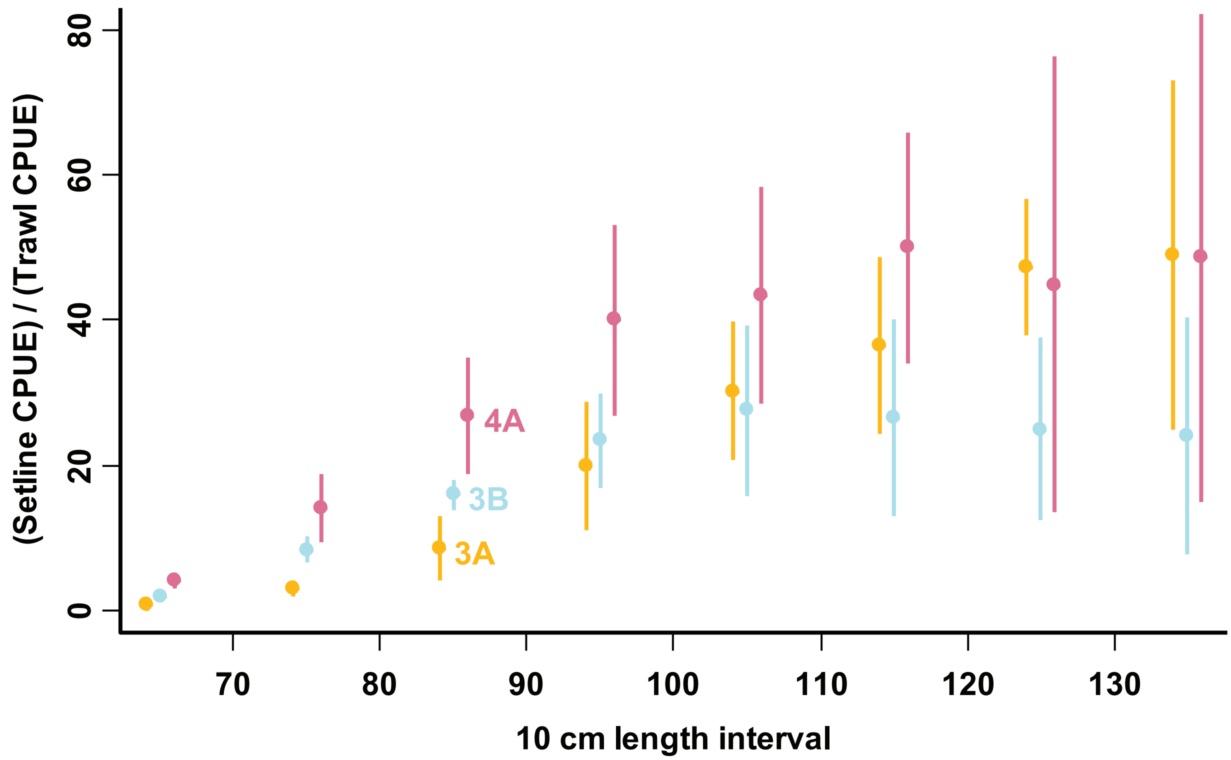

Figure 1. Ratios of setline to trawl survey CPUE at length by area.

Assessment of the Pacific halibut stock at the end of 2007

William G. Clark and Steven R. Hare

Abstract

As in 2006, the stock assessment was done by fitting the assessment model to a coastwide dataset to estimate total biomass, and then apportioning the total among regulatory areas in accordance with survey estimates of relative abundance. Coastwide exploitable biomass in 2008 is estimated to be 361 million pounds, down from the 414 million estimated last year. About half of the decrease is due to a change in the parameterization of survey catchability in the model, and the other half to lower commercial and survey catch rates in 2007. Total CEY is 69 million pounds.

Introduction

Each year the International Pacific Halibut Commission (IPHC) staff assesses the abundance and potential yield of Pacific halibut using all available data from the commercial fishery and scientific surveys (Appendix A). A biological target level for total removals from each regulatory area is calculated by applying a fixed harvest rate to the estimate of exploitable biomass in that area. This target level is called the "constant exploitation yield" or CEY for that area in the coming year. The corresponding target level for catches in directed fisheries subject to allocation is called the fishery CEY. It comprises the commercial setline catch in all areas plus the sport catch in Areas 2A and 2B. It is calculated by subtracting from the total CEY an estimate of all unallocated removals—bycatch of legal-sized fish, wastage of legal-sized fish in the halibut fishery, fish taken for personal use, and sport catch except in Areas 2A and 2B. Staff recommendations for catch limits in each area are based on the estimates of fishery CEY but may be higher or lower depending on a number of statistical, biological, and policy considerations. Similarly, the Commission's final quota decisions are based on the staff's recommendations but may be higher or lower.

For many years the staff assessed the stock in each regulatory area by fitting a model to the data from that area (Appendix B). This procedure relied on the assumption that the stock of fish of catchable size in each area was closed, meaning that net migration was negligible. A growing body of evidence from both the assessments (Clark and Hare 2007a) and the ongoing mark-recapture experiment (Webster and Clark 2007) shows that there is probably a continuing eastward net migration of catchable fish from the western Gulf of Alaska (Areas 3B and 4) to the eastern side (Area 2). The effect of this migration on the closed-area stock assessments was to produce underestimates of abundance in the western areas and overestimates in the eastern areas. To some extent this has almost certainly been the case for some time, meaning that exploitation rates were well above the target level in Area 2 and a disproportionate share of the catches have been taken from there.

In order to obtain an unbiased estimate of the coastwide stock in the 2006 assessment, the staff built a coastwide data set and fitted the model to it. Exploitable biomass in each regulatory area was estimated by apportioning the total in proportion to an estimate of stock distribution derived from the setline survey catch rates (CPUE). Specifically, an index of abundance in each area was calculated by multiplying survey CPUE (running 3-year average) by total bottom area between 0 and 300 fm. The logic of this index is that survey CPUE can be regarded as an index of density, so multiplying it by bottom area gives a quantity proportional to total abundance. The estimated proportion in each area is then the index value for that area divided by the sum of the index values. This year's assessment uses the same procedure.

Description of the assessment model

The IPHC assessment model is age- and sex-structured. Commercial and survey selectivity are both estimated as piecewise linear functions of observed mean length at age/sex in survey catches. (There is a 32" minimum size limit in the commercial fishery.) Commercial catchability is normally allowed to vary from year to year with a penalty of 0.03 on log differences. Survey catchability is normally held constant, although some variation was allowed in both this year's and last year's production fits. The model is fitted to commercial and survey catch at age and CPUE. Clark and Hare (2006) provide a full account of model structure and fitting procedures.

The closed-area and coastwide model fits differ in parameterization and likelihood. Some of the closed-area data sets are quite noisy, so the closed-area version is more parsimonious and it is weighted. Specifically, the catchability, selectivity and natural mortality parameters are all unisex; the estimated selectivity schedules are strongly smoothed; the model is fitted only to total CPUE (rather than CPUE at age/sex); and a heavy weight is placed on the CPUE data series to assure satisfactory agreement. The coastwide data are not noisy, so the coastwide version of the model can have sex-specific parameters, weaker selectivity smoothing, and neutral data weighting. It is fitted to CPUE at age/sex as well as total CPUE.

Alternative model fits

In the 2006 coastwide assessment (Clark and Hare 2007b), estimated survey catchability was allowed to vary somewhat because it was found that actual survey catchability had varied substantially. This was shown by model fits in which present abundance was fixed at a range of levels by fixing the terminal fishing mortality rate as in a virtual population analysis (VPA) and then estimating survey catchability as a free parameter in each year (Fig. 1). These fits showed that survey catchability happened to be high in the first year of the data (1997) and low in the last year (2006), resulting in a spurious appearance of a decline in abundance. To neutralize that feature, survey catchability was estimated independently for the first and last years, which effectively meant disregarding those data points and estimating a constant survey catchability from the remaining data (1998-2005).

In this year's assessment some other ways of dealing with variable survey catchability were considered. The candidate models were:

(i) Vanilla: the conventional model, with constant survey catchability in all years.

(ii) HiLoSQ: last year's production model, with three values of survey catchability estimated (1997, 1998-2005, 2006-2007).

(iii) WobbleSQ: survey catchability estimated for each year, but with a penalty of 0.05 on log differences. This is similar to the treatment of commercial catchability.

(iv) TrendlessSQ: same as WobbleSQ, but with the additional requirement that a regression of estimated survey catchability on year have zero slope. This means that survey catchability was allowed to vary but not to show any trend over time.

Table 1 shows features of the candidate model fits and some others. WobbleSQ has the lowest AIC score, but TrendlessSQ is nearly as good, and we think it is appropriate to disallow trends in survey catchability over time, so that is our chosen production model.

The last two fits in Table 1 show the effect of commercial CPUE on the biomass estimate. "No commercial CPUE" is a fit in which commercial CPUE is disregarded, and "CAGEAN" is a fit in which commercial catchability is held constant, so that commercial and survey CPUE are given equal influence. Evidently commercial CPUE tends to increase the biomass estimate, but not greatly.

Effect of the 2007 data on abundance estimates

Coastwide commercial and survey CPUE both declined by 5-10% from 2006 to 2007 (Fig. 2; Appendix A tables A2 and A3). As a result the 2007 coastwide and closed-area model fits mostly revise downward the estimates of abundance at the beginning of 2007 made in the 2006 assessment (Table 2). At the same time the 2007 fits show an increase in abundance between the beginning of 2007 and the beginning of 2008, so last year's estimates of 2007 biomass and this year's estimates of 2008 biomass are not very different in most cases. Exceptions are Areas 2C and 4A where the closed-area estimates decrease significantly.

The coastwide estimate of exploitable biomass in 2008 is 361 M lb compared with 414 last year. About half of this difference is due to the change from the HiLoSQ to the TrendlessSQ model fit. The HiLoSQ biomass estimate in 2008 is 386 M lb.

Area-specific biomass and CEY estimates

Area-specific estimates of biomass are calculated by survey apportionment as they were last year, with the difference that this year a depth-stratified mean survey CPUE has been used, which results in about a 40% increase in the Area 2A apportionment, about a 5% decrease in the Area 3A apportionment, and very small increases in most other apportionments. The area-specific estimates from last year's and this year's coastwide and closed-area assessments are shown in Tables 3 and 4.

The staff believes that survey apportionment is the most objective and consistent method of estimating the biomass distribution among areas and therefore the best distribution of total CEY, if the aim is proportional harvest. A disproportionate share of the harvest has been taken from Area 2 for decades, so some level of disproportionality was clearly sustainable by the stock with the exploitation pattern that prevailed during that period. Increasing catches from the western portion of the stock in the last decade have altered the exploitation pattern, so the historical high levels of removals from Area 2 may no longer be sustainable. Alternative CEY apportionments under a variety of rules are shown for information in Table 6. The staff does not advocate any of them and would in fact oppose some, such as apportionment on the basis of bottom area alone or an index incorporating commercial CPUE.

Evaluation of the assessment

Quality of fits

The assessment model fits the coastwide data very well. (That is not true of some of the closed-area data sets.) The series of total survey and commercial CPUE are predicted closely (Fig.

3, bottom panels), and so are the commercial catch and survey CPUE at age/sex (Figs. 4a and 4b).

Retrospective performance

Each year's model fit estimates the abundance and other parameters for all years in the data series. One hopes that the present assessment will closely match the biomass trajectory estimated by the previous year's assessment. To the extent that it does not, the assessment is said to have poor retrospective performance.

Our assessment has not tracked very well for the last few years. Each year the assessment has revised downward the previous year's biomass estimates (Fig. 5), meaning that biomass was overestimated then and may be overestimated now if the cause of the retrospective problem lies somewhere within the model. There is some precedent for that; the assessment models in use in the mid 1990s and the early 2000s showed strong retrospective patterns that turned out to be the result of misspecified selectivity (age- rather than length-based). There is also the possibility that the retrospective pattern is caused in some way by the external estimation of the sex composition of the commercial catch, or by the internal prediction of surface age compositions prior to 2002 through the application of an age misclassification matrix (Clark and Hare 2006).

Problems of this sort with the assessment machinery would manifest themselves as systematic revisions of the estimated relative strength of the year-classes present in the stock. That was true of the retrospective patterns caused by the misspecification of selectivity in the past: incoming year-classes would at first be estimated as weak because catch rates were low, but the real reason was low selectivity rather than low abundance. When they were later caught in large numbers, the estimates of relative year-class strength increased.

We can check for patterns of this sort by doing a blind projection of the assessment from, say, 2004 to 2007. This means using the estimates of year-class strength and other parameters from the 2004 assessment and projecting forward to 2007 without benefit of the 2005-2007 data (except for the total catch in number in each year, which determines the annual fishing mortality rate). If there were some problem with the model, the projected age compositions of the survey and commercial catches would differ systematically from the predictions of the 2007 assessment incorporating the 2005-2007 data. But they do not; the two sets of predicted age compositions are nearly the same (Fig. 6a). This is not surprising, given the simplicity of the model and the very good fits to the data.

What the projection from 2004 fails to predict is the commercial and survey CPUE in 2005-2007 (Fig. 6b). Given the estimates of year-class strength and catchability in 2004, the blind projection shows CPUE bottoming out in 2005 and increasing thereafter. In actuality both declined in 2006 and again in 2007, with the result that the present abundances of all of the year-classes in the stock were revised downward proportionally in the subsequent assessments. So this is a retrospective pattern caused by the data, not by the model.

To some extent the pattern results from the decline in survey catchability mentioned above. VPA-like fits in 2007 show that survey catchability declined every year from 2005 through 2007, by some 20% in total. This is by no means unprecedented, but the run of three declines in a row inevitably affects the biomass estimates. This year's production model ("Trendless") is less affected than a conventional model ("Vanilla") because it allows survey catchability to vary from year to year, but it is affected.

Estimates of uncertainty

There are a number of ways of estimating the uncertainty associated with a given model fit and biomass estimate. They are all unsatisfactory in that they are conditioned on the correctness of the model, and in fact it is the choice of one model rather than another that is the major source of uncertainty in assessments. This is well illustrated by the difference in area-specific biomass estimates between the coastwide and closed-area fits of the IPHC model.

Figure 7 shows probability distributions of the 2008 exploitable biomass obtained in various ways. The Hessian-based estimate of standard deviation is about 20 M lb, and a normal distribution with this amount of dispersion closely approximates a calculated likelihood profile. A straightforward measure of uncertainty is the spread of biomass estimates among plausible models. All of the fits in Table 1 are at least plausible, and they range from 320 to 400 M lb, similar to the Hessian-based normal approximation.

Treatment of process error

The likelihood used in fitting the model is the MULTIFAN scheme developed by Fournier et al. (1990). All errors are treated as being normally distributed, so the likelihood is a sum of squared deviations, each weighted by the inverse of a scaled variance. The variances are the external estimates of sampling variance of each observation, and the scalers are just the root mean squared errors associated with each data type in unscaled fits. This amounts to a one-step reweighting of the data. It succeeds in producing distributions of residuals that are very close to standard normals. The scalers are mostly in the range 4-9, meaning that sampling variance accounts for only a small fraction of the total error variance. The remainder is process error, the result of model misspecification or parameter variation

While the MULTIFAN procedure is clearly effective in standardizing the variances in the halibut assessment, it is somewhat puzzling that process error can be successfully treated as a multiple of sampling error. They arise from different sources and there is really no reason to expect them to be related. One suggestion made during an external review in 2007 was that we consider an additive rather than a multiplicative model of process error. The multiplicative model is $\sigma_p^2 = \left(\tau^2 - 1\right) \cdot \sigma_s^2$, where σ_p^2 is process variance, σ_s^2 is sampling variance, and τ^2 is a scaler. Total variance σ_t^2 is then given by $\sigma_t^2 = \tau^2 \cdot \sigma_s^2$. The additive model is $\sigma_t^2 = \sigma_s^2 + \sigma_p^2$ where σ_p^2 is process error. The suggestion was to estimate a process coefficient of variation (CV) for each data type, so $\sigma_p^2 = \delta^2 \cdot y^2$ where y is the observed value and δ is the CV.

The amount of process error associated with each data point can be estimated as the squared deviation (in an unscaled fit) minus the estimated sampling variance. If the multiplicative model is appropriate, process error should increase with sampling variance, and it does (Fig. 8a). If the additive model is appropriate, process error should increase with the square of the observed value, and it does (Fig. 8b). The reason that both models are appropriate is that most of the observations (commercial and survey catch and CPUE at age/sex) have multinomial sampling variances, so the sampling variances are proportional to the expected values. So while equally appropriate, the additive model would not improve on the multiplicative model.

Use of PIT tag estimates of commercial selectivity in the assessment

Estimates of fishing mortality from the ongoing PIT tag experiment (Webster 2008) are so different from the stock assessment as to be simply incredible, but that is not true of the selectivity estimates. Even when mark-recapture data are not usable for estimating fishing mortality or abundance or migration rates, they can still provide useful estimates of selectivity (Myers and Hoenig 1997, Clark and Kaimmer 2006).

In the stock assessment, commercial selectivity is required to reach 100% at a length of 120 cm and remain there (i.e., commercial selectivity is asymptotic). In model fits, commercial selectivity increases gradually between 80 and 120 cm. At 100 cm it is estimated to be 0.56. The PIT tag data show full commercial selection occurring at a smaller size than the assessment. When a coastwide commercial selectivity is estimated freely from the PIT tag data, it reaches 100% at 100 cm and stays close to that level thereafter (Ray Webster, IPHC, pers. comm.).

The assessment can be made to conform to the PIT tag results by requiring full commercial selection at 100 cm. When that is done, the fit is much worse (AIC = 850 vs 790 for the production model). The exploitable biomass estimate is nearly the same (373 M lb vs 361).

References

Clark, W. G., and Hare, S. R. 2006. Assessment and management of Pacific halibut: data, methods, and policy. Int. Pac. Halibut Comm. Sci. Rep. 83.

Clark, W. G., and Hare, S. R. 2007a. Motivation and plan for a coastwide stock assessment. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006 83-96.

Clark, W. G., and Hare, S. R. 2007b. Assessment of the Pacific halibut stock at the end of 2006. Int. Pac. Halibut Comm. Report of Research and Assessment Activities 2006:97-128.

Clark, W. G., and Kaimmer, S. M. 2006. Estimates of commercial longline selectivity for Pacific halibut (*Hippoglossus stenolepis*) from multiple marking experiments. Fish. Bull. 104:465-467.

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 from southern bluefin tuna (*Thunnus maccoyi*). Can. J. Fish. Aquat. Sci. 47:301-317.

Myers, R. A., and Hoenig, J. M. 1997. Direct estimates of gear selectivity from multiple tagging experiments. Can. J. Fish. Aquat. Sci. 54:1-9.

Webster, R. 2008. Analysis of PIT tag recoveries through 2007. Int. Pac. Halibut Comm. Report of Assessmnent and Research Activities 2007 259-274.

Webster, R., and Clark, W. G. 2007. Analysis of PIT tag reoveries through 2006. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:129-138.

Table 1. Alternative coastwide model fits. The first two are coastwide fits that have the same parameterization as the closed-area fits.

Model	Number of parameters	Deviance	AIC	Exploitable biomass
Closed-area parameters Closed-area likelihood	121	NA	NA	321
Closed-area parameters Coastwide likelihood	121	716	958	341
Vanilla	136	524	796	337
WobbleSQ	155	479	789	338
HiLoSQ	138	520	796	386
TrendlessSQ	155	480	790	361
No commercial CPUE	145	504	794	344
CAGEAN	134	553	821	400

Table 2. Effect of the 2007 data on closed-area and coastwide abundance estimates.

Area	2007 ebio 2006 assessment Data as of 11/06	2007 ebio 2006 assessment Data as of 11/07	2007 ebio 2007 assessment Data as of 11/07	2008 ebio 2007 assessment Data as of 11/07
Closed-area assessments:				
2A	4.9	5.1	4.0	4.6
2B	39	41	33	37
2C	57	55	45	49
3A	174 [1]	170	169	169
3B	52	53	47	54
4A	17	14	11	11
4B	10	12	15	14
2A-4B sum	354	350	324	339
4CDE	58	52	52	52
Total	412	402	376	391
Coastwide assessment:				
2A-4B sum (90% of total)	339	333	297	325
4CDE	38	37	33	36
Total	377	370	330	361

Notes:

[1] Recalculated to be consistent with present treatment of Area 3A survey CPUE (full-area CPUE = 81% of partial-area CPUE rather than 75%). Value reported last year was 186.

Table 3. Estimates of 2007 exploitable biomass and CEY from the 2006 assessment (2006 RARA, p. 107).

	Area 2A	Area 2B	Area 2C	Area 3A	Area 3B	Area 4A	Area 4B	Area 4CDE	Total
Coastwide assessment [1]									
2007 exploitable biomass	3.7	27	33	176	86	29	19	41	414
Proportion of total	0.009	0.065	0.080	0.423	0.208	0.069	0.045	0.101	1.000
Target harvest rate	0.25	0.25	0.25	0.20	0.20	0.20	0.15	0.15	~0.20
Total CEY	0.93	6.75	8.25	35.20	17.20	5.80	2.85	6.15	83.13
Other removals [2]	0.27	0.53	3.79	7.89	0.43	0.57	0.29	2.30	16.07
2007 fishery CEY [2]	0.66	6.22	4.46	27.31	16.77	5.23	2.56	3.85	67.06
Area assessments [1]									
2007 exploitable biomass	4.9	39	57	186	52	17	10	50	416
Proportion of total	0.012	0.094	0.137	0.447	0.125	0.041	0.024	0.120	1.000
Target harvest rate	0.20	0.20	0.20	0.20	0.20	0.20	0.15	0.15	~0.20
Total CEY	1.00	7.80	11.40	37.20	10.40	3.40	1.50	7.50	80.20
Other removals [2]	0.27	0.53	3.79	7.89	0.43	0.57	0.29	2.30	16.07
2007 fishery CEY [2]	0.73	7.27	7.61	29.31	9.97	2.83	1.21	5.20	64.13
2007 catch limit [3]	1.34	11.47	8.51	26.20	9.22	2.89	1.44	4.10	65.17

Notes:

[1] "Coastwide assessment" refers to the coastwide model fit with survey apportionment of the total biomass estimate among regulatory areas. "Area assessments" are the closed-area model fits.

[2] "Other removals" comprise legal-sized wastage, legal-sized bycatch, personal use, and in most areas sport catch. In Areas 2A and 2B sport catch is included in fishery CEY rather than in other removals.

[3] "Catch limit" includes sport as well as commercial catch in Areas 2A and 2B.

Table 4. Estimates of 2008 exploitable biomass and CEY from the 2007 assessment.

	Area 2A	Area 2B	Area 2C	Area 3A	Area 3B	Area 4A	Area 4B	Area 4CDE	Total
Coastwide assessment [1]									
2008 exploitable biomass	4.7	25.6	32.5	144.8	74.0	21.3	20.2	37.9	361
Proportion of total	0.013	0.071	0.090	0.401	0.205	0.059	0.056	0.105	1.000
Target harvest rate	0.20	0.20	0.20	0.20	0.20	0.20	0.15	0.15	<0.20
Total CEY	0.94	5.12	6.50	28.96	14.80	4.26	3.03	5.69	69.30
Other removals [2]	0.29	0.47	2.59[3]	6.71[3]	0.53	0.75	0.33	2.01	13.68
2008 fishery CEY [2]	0.65	4.65	3.92	22.25	14.27	3.51	2.71	3.68	55.62
Area assessments [1]									
2008 exploitable biomass	4.6	37	49	169	54	11	14	52	391
Proportion of total	0.012	0.095	0.125	0.432	0.138	0.028	0.036	0.133	0.999
Target harvest rate	0.20	0.20	0.20	0.20	0.20	0.20	0.15	0.15	<0.20
Total CEY	0.92	7.40	9.80	33.80	10.80	2.20	2.10	7.80	74.82
Other removals [2]	0.29	0.47	2.59[3]	6.71[3]	0.53	0.75	0.33	2.01	13.68
2008 fishery CEY [2]	0.63	6.93	7.21	27.09	10.27	1.45	1.77	5.79	61.14

Notes:

[1] "Coastwide assessment" refers to the coastwide model fit with survey apportionment of the total biomass estimate among regulatory areas. "Area assessments" are the closed-area model fits.

[2] "Other removals" comprise legal-sized wastage, legal-sized bycatch, personal use, and in most areas sport catch. In Areas 2A and 2B sport catch is included in fishery CEY rather than in other removals.

[3] The sport catch component in these figures is the adopted guideline harvest level (GHL), which is lower than the actual 2007 catch in Area 2C and higher in Aea 3A.

Table 5. Other removals in detail. Sport catch figures for Areas 2C and 3A are actual catches not GHL levels as in Table 4.

	Area 2A	Area 2B	Area 2C	Area 3A	Area 3B	Area 4A	Area 4B	Area 4CDE	Total
Sport catch	0.52	1.75	2.55	5.05	0.01	0.05	0.00	0.00	9.93
Legal-sized bycatch	0.25	0.15	0.21	0.99	0.45	0.66	0.32	1.90	4.93
Pesonal use	0.04	0.30	0.58	0.38	0.05	0.03	0.00	0.11	1.49
Legal-sized wastage	0.00	0.02	0.02	0.05	0.02	0.01	0.01	0.00	0.13
Total	0.81	2.22	3.36	6.47	0.53	0.75	0.33	2.01	16.48
Total excl.sport catch in Areas 2A and 2B	0.29	0.47	3.36	6.47	0.53	0.75	0.33	2.01	14.21
Sublegal discard mortality (shown for information; not taken off total CEY)	0.02	0.44	0.27	0.92	0.42	0.13	0.02	0.07	2.29

Table 6. Shares of total CEY by area according to various apportionment rules.

Rule	Area 2A	Area 2B	Area 2C	Area 3A	Area 3B	Area 4A	Area 4B	Area 4CDE	Total
Survey apportionment (CPUE x bottom area)	0.013	0.071	0.090	0.401	0.205	0.059	0.056	0.105	1.000
2008 exploitable biomass from 2007 closed-area assessments	0.012	0.095	0.125	0.432	0.138	0.028	0.036	0.133	0.999
Historical recruitment from 2007 closed-area assessments (1987-1996)	0.02?	0.107	0.098	0.451	0.161	0.046	0.018	0.10?	1.001
Share of total catch (1990-2007)	0.017	0.144	0.140	0.366	0.142	0.065	0.035	0.091	1.000
Share of bottom 0-300 fm (excl. EBS shelf outside 4C)	0.066	0.160	0.082	0.256	0.154	0.094	0.078	0.111	1.001
Commercial apportionment (CPUE x bottom area)	0.035	0.113	0.055	0.401	0.162	0.088	0.066	0.080	1.000

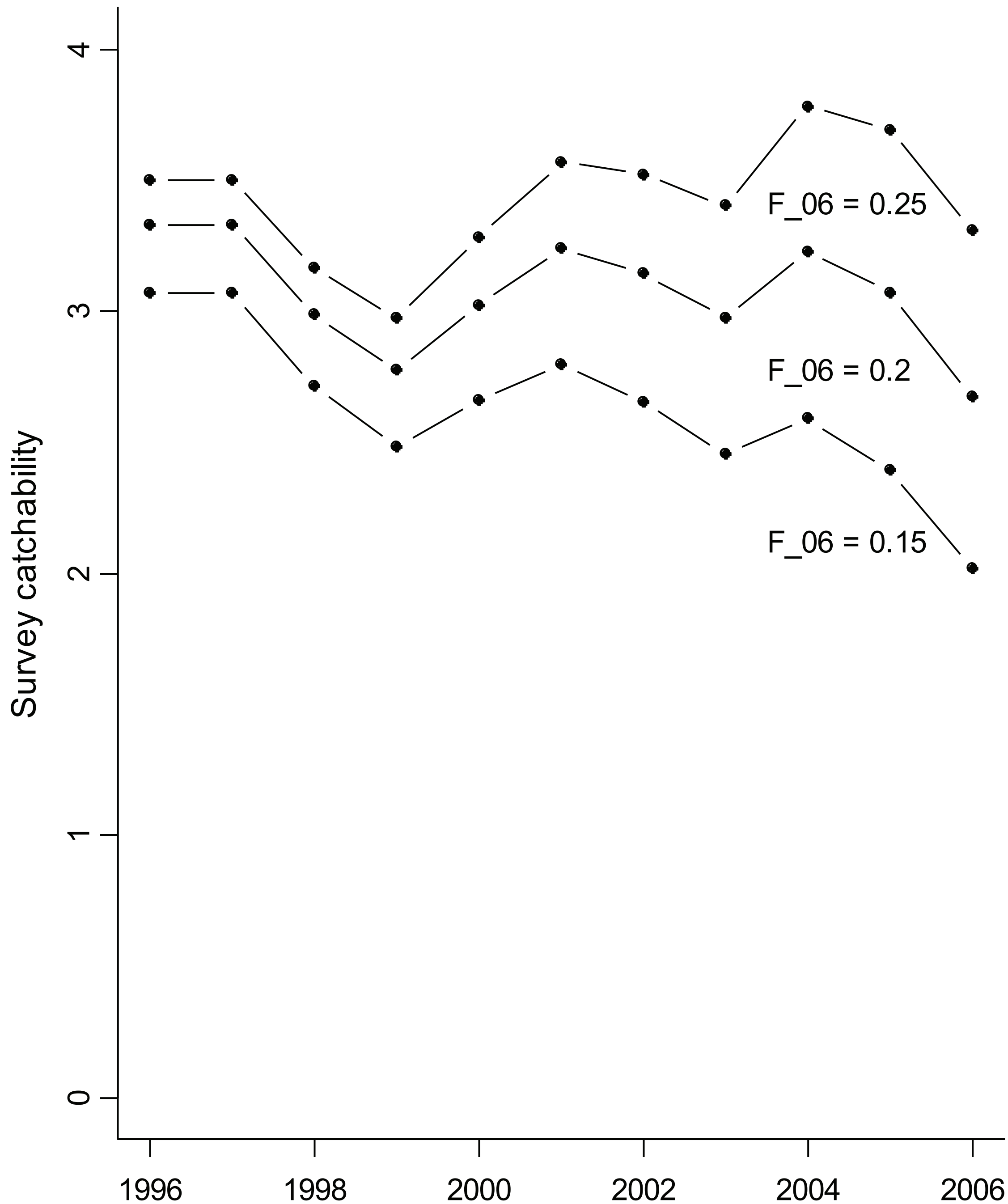

Figure 1. Calculated values of survey catchability in VPA-like fits of the model in the 2006 assessment. The labels refer to the value of the fixed terminal fishing mortality rate; e.g. "F_06 = 0.2" means that the fishing mortality rate in 2006 was set to 0.20.

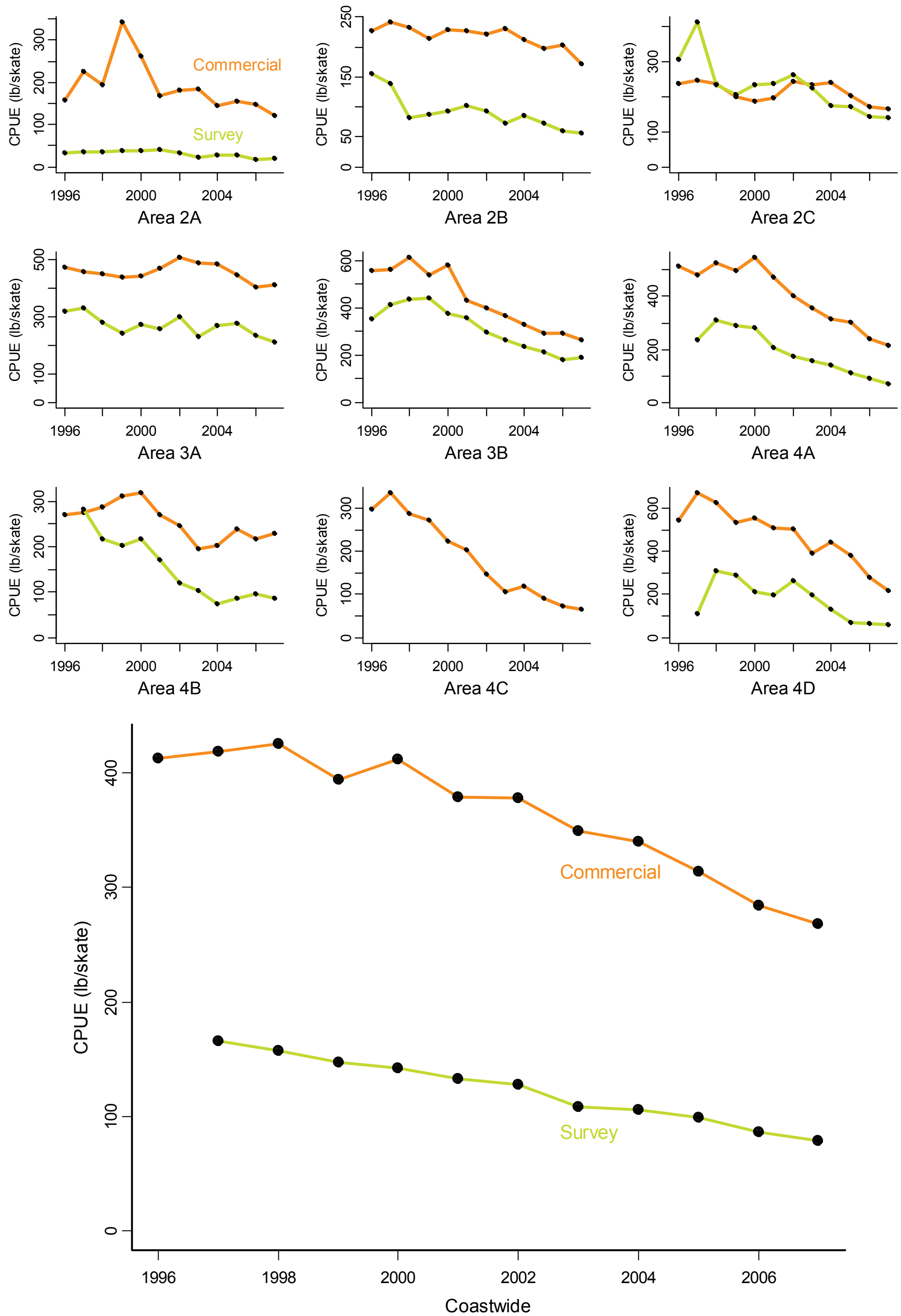

Figure 2. Commercial and survey CPUE by area (above) and coastwide (below).

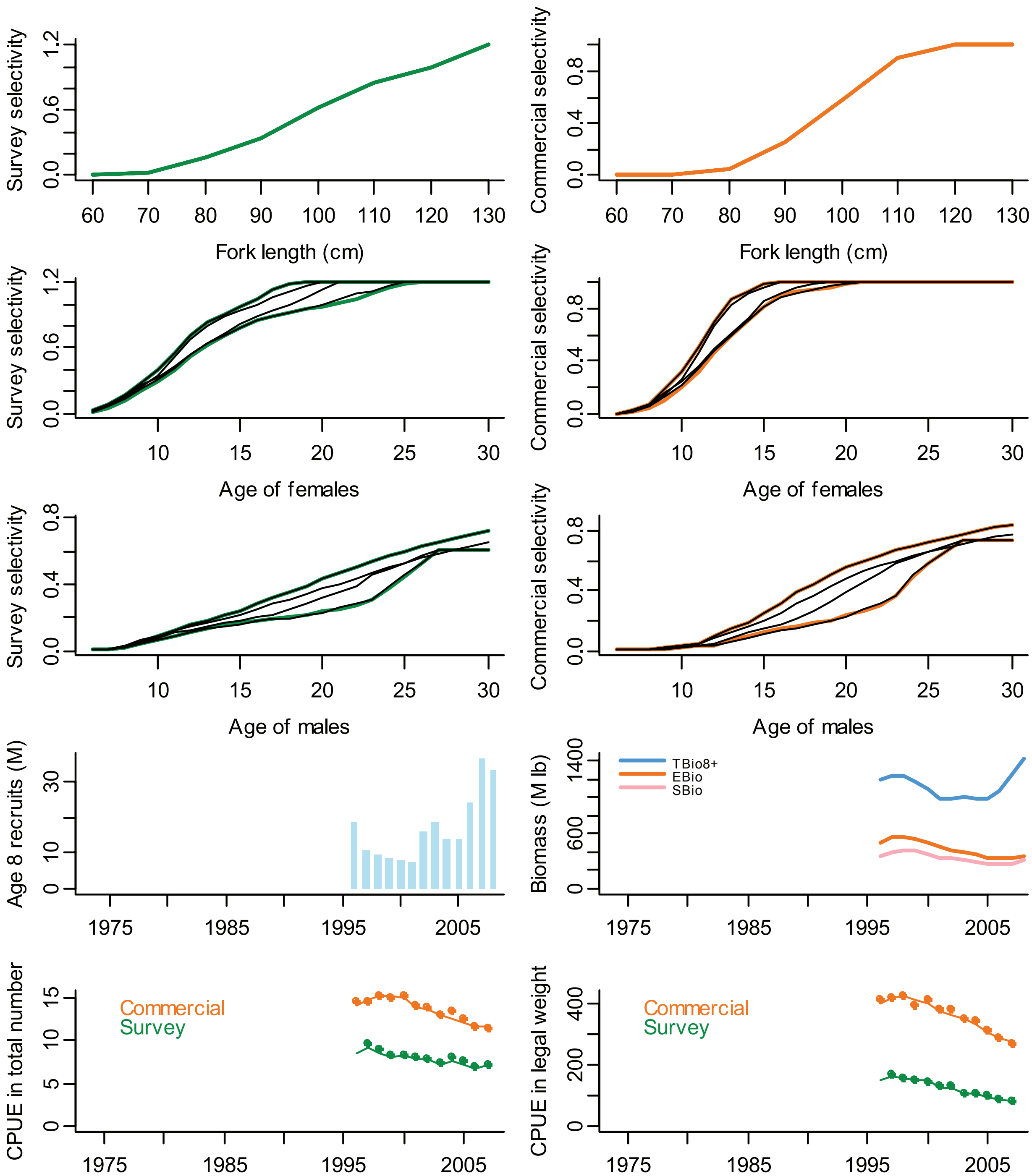

Figure 3. Features of the 2007 coastwide assessment. Age-specific selectivities are plotted for every third year plus the last.

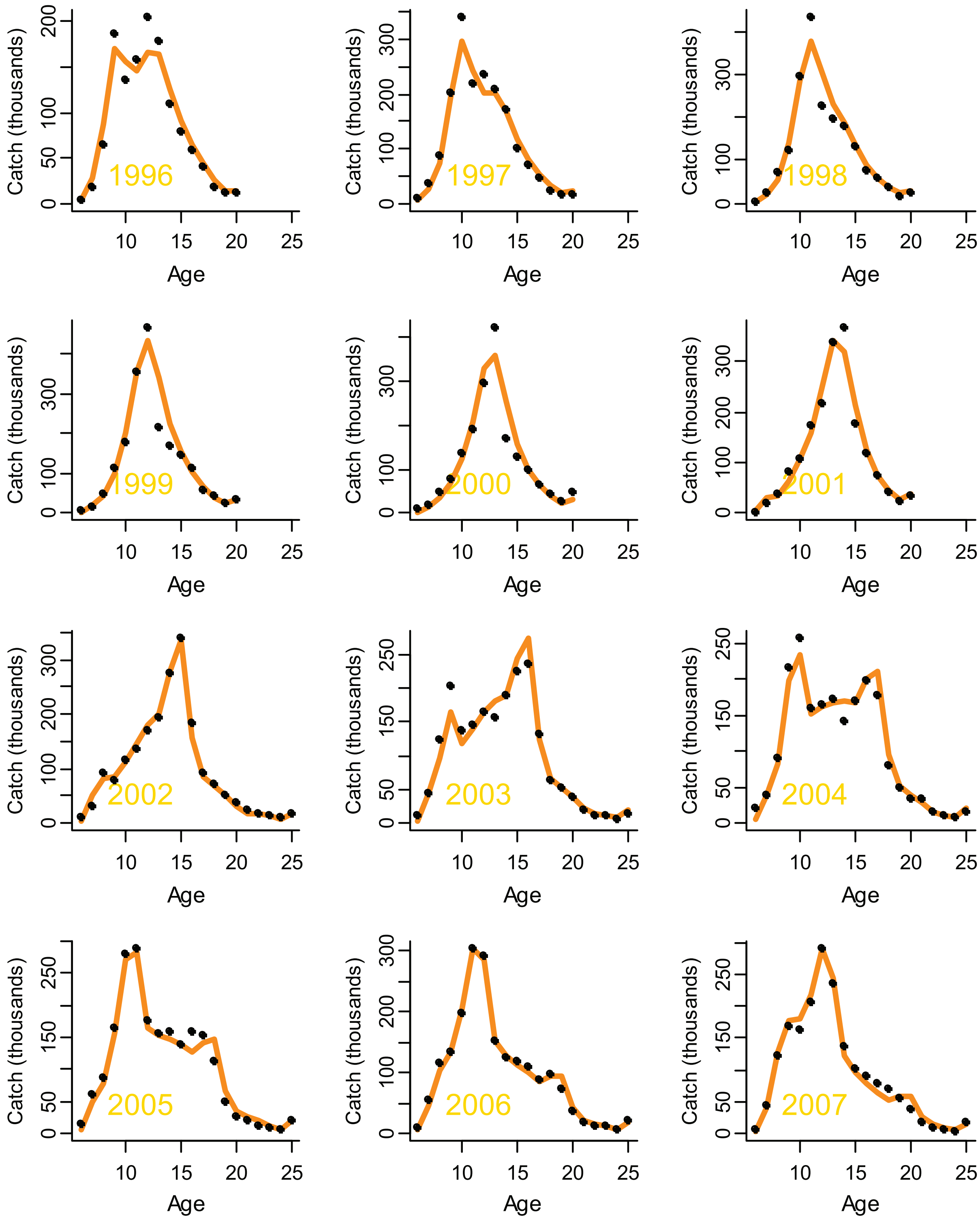

Figure 4a. Observed (points) and predicted (lines) commercial catch at age of females in the 2007 coastwide model fit.

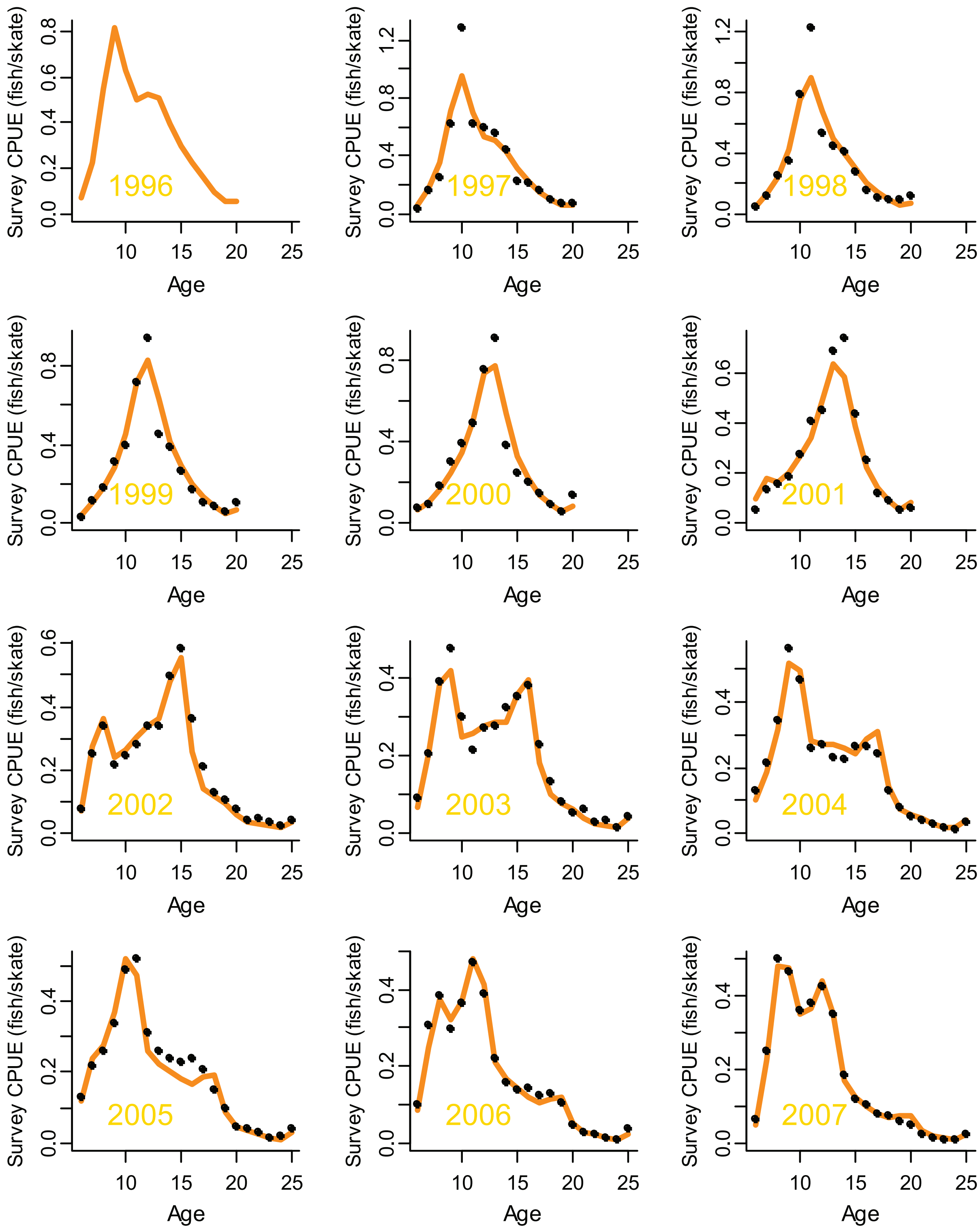

Figure 4b. Observed (points) and predicted (lines) survey CPUE at age of females in the 2007 coastwide model fit.

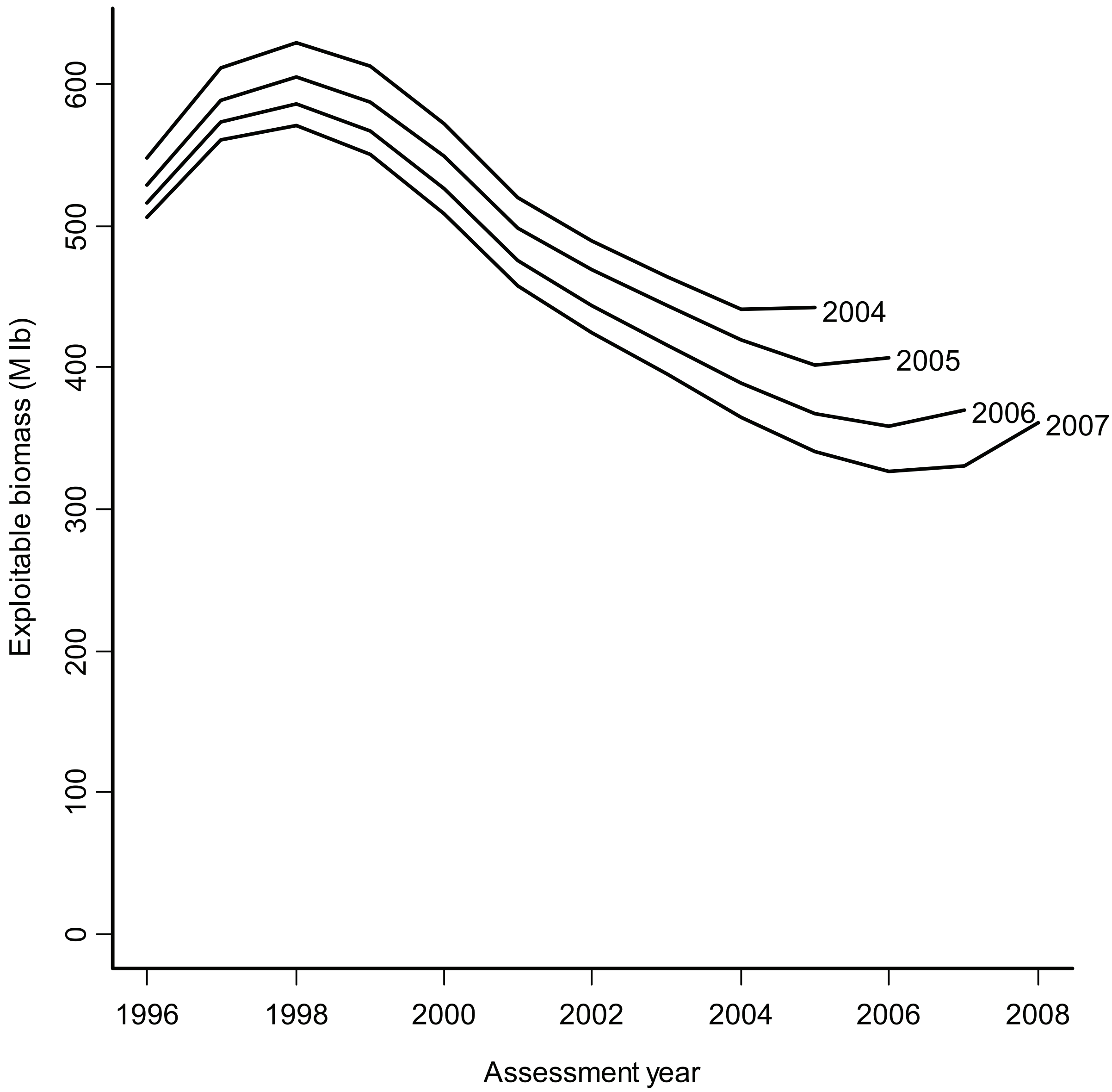

Figure. 5. Retrospective performance of the assessment. Each line is the biomass trajectory estimated by the model fitted to data from 1996 through the labeled last year.

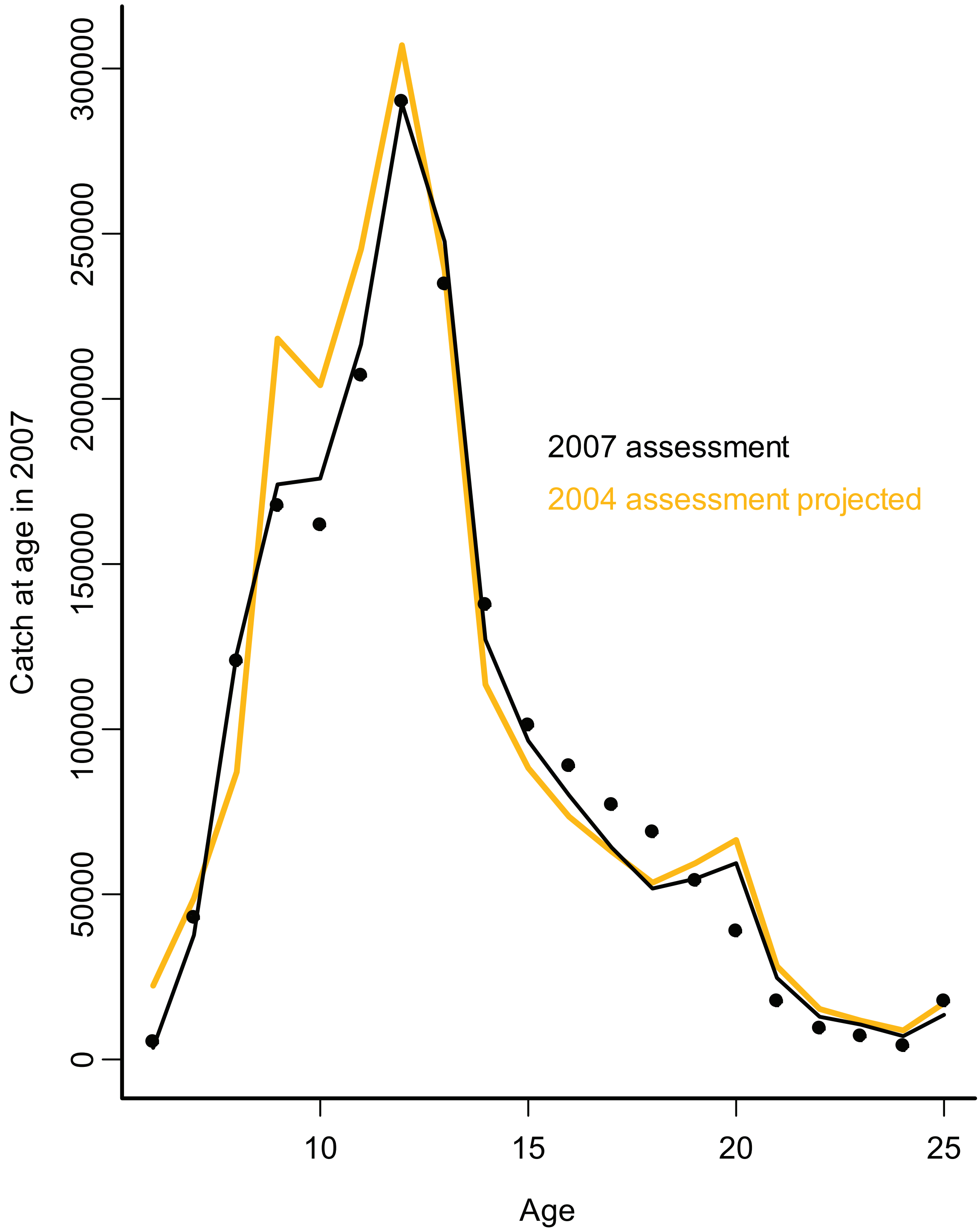

Figure 6a. Observed commercial catch at age of females in 2007 (points) and predicted catch at age from the 2007 assessment and from a blind projection of the 2004 assessment.

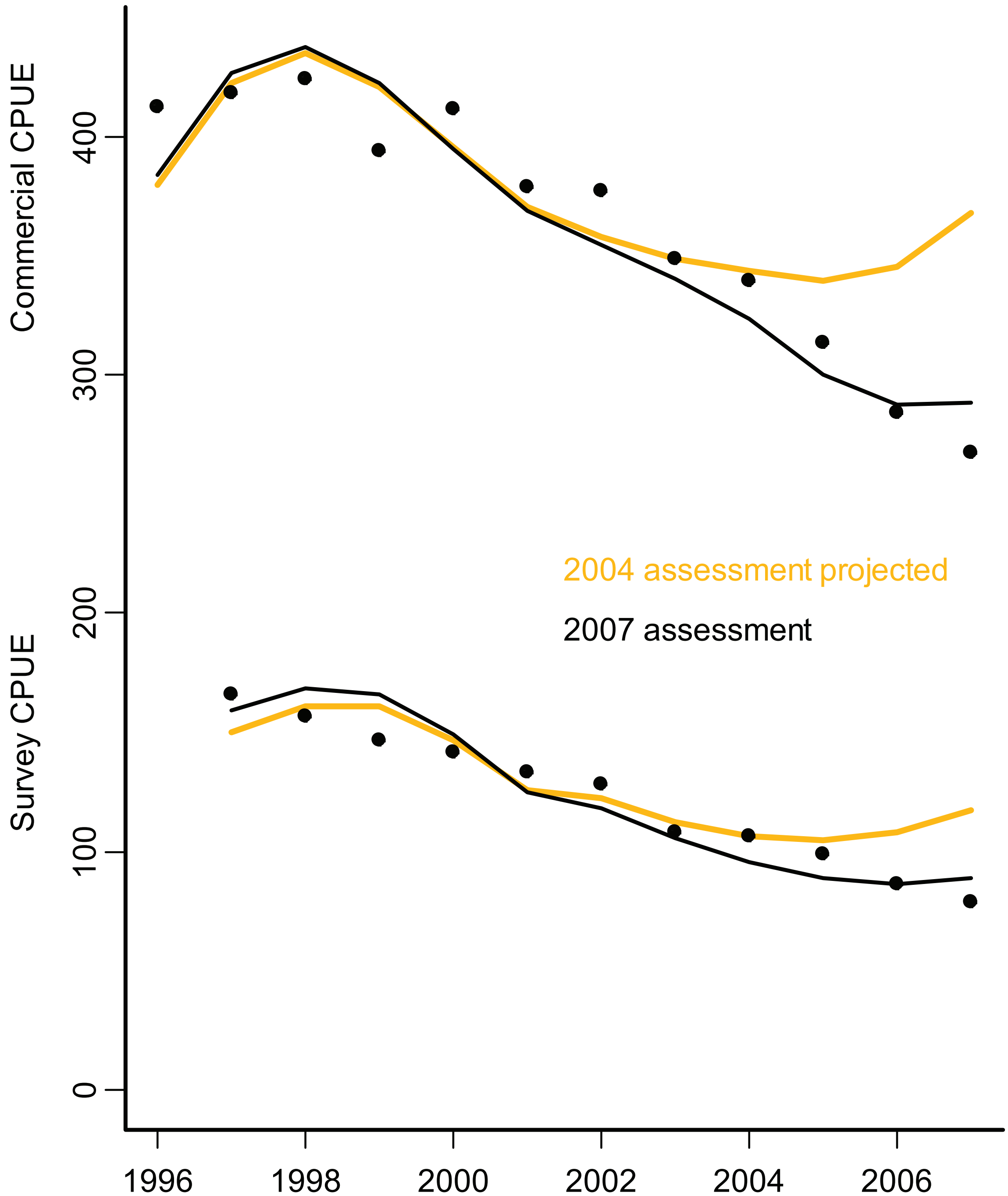

Figure 6b. Points are observed commercial (above) and survey (below) CPUE. Lines are predicted values from the 2007 assessment and a blind projection of the 2004 assessment.

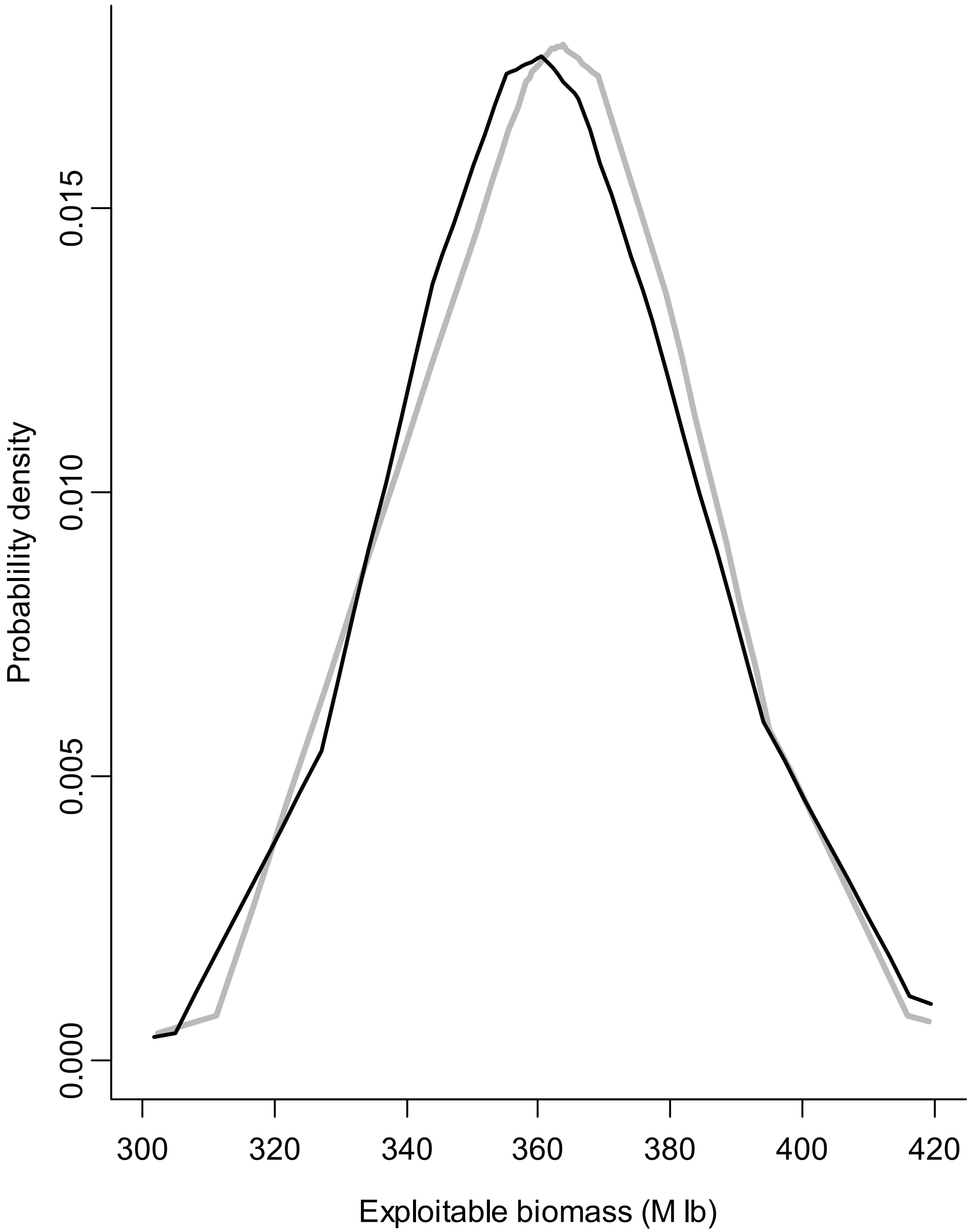

Figure 7. Estimates of uncertainty in the estimate of 2008 exploitable biomass: normal approximation based on the Hessian (gray line) and calculated likelihood profile (black line).

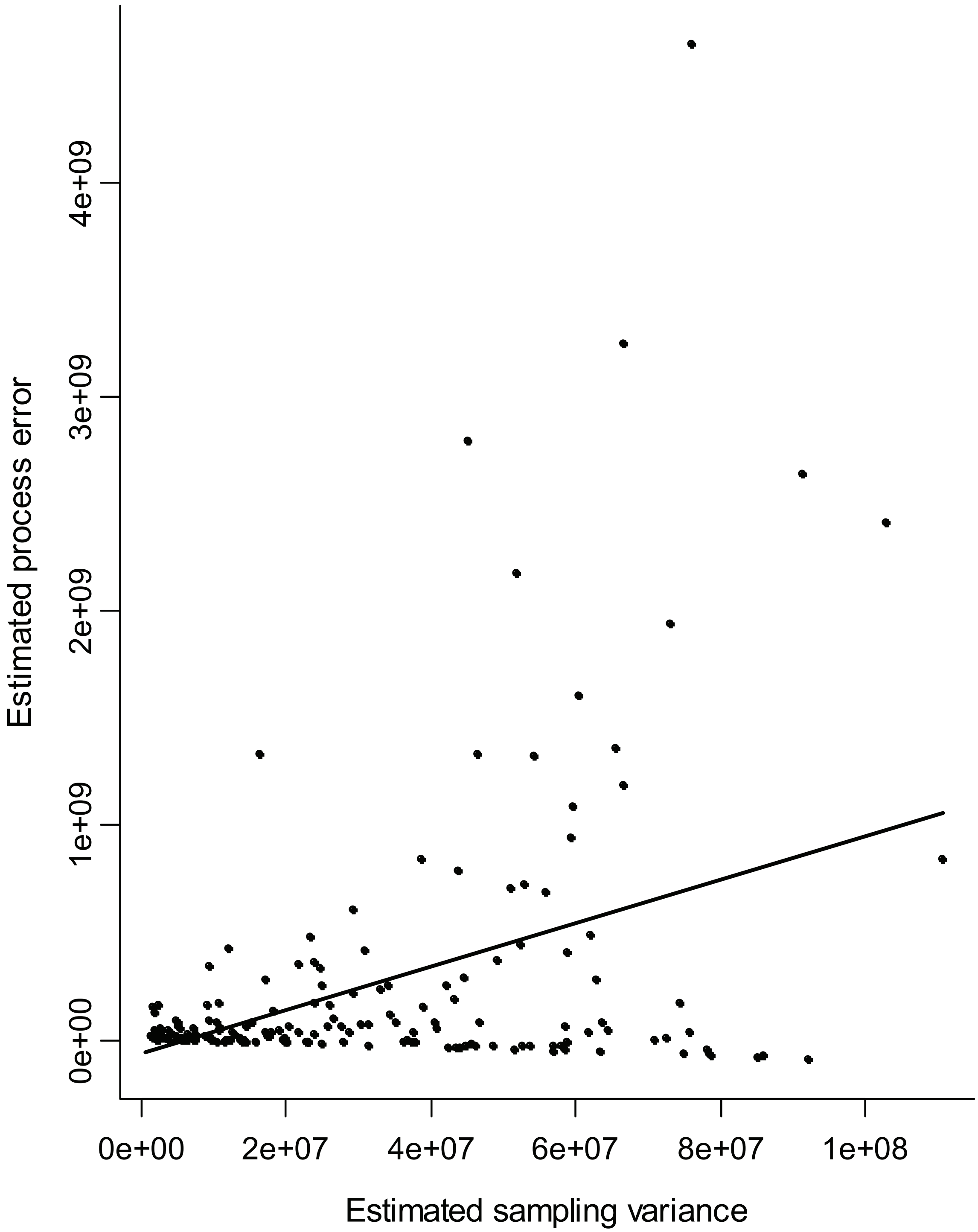

Figure 8a. Estimated process error (squared deviation minus estimated sampling variance) plotted against estimated sampling variance of female catch at age.

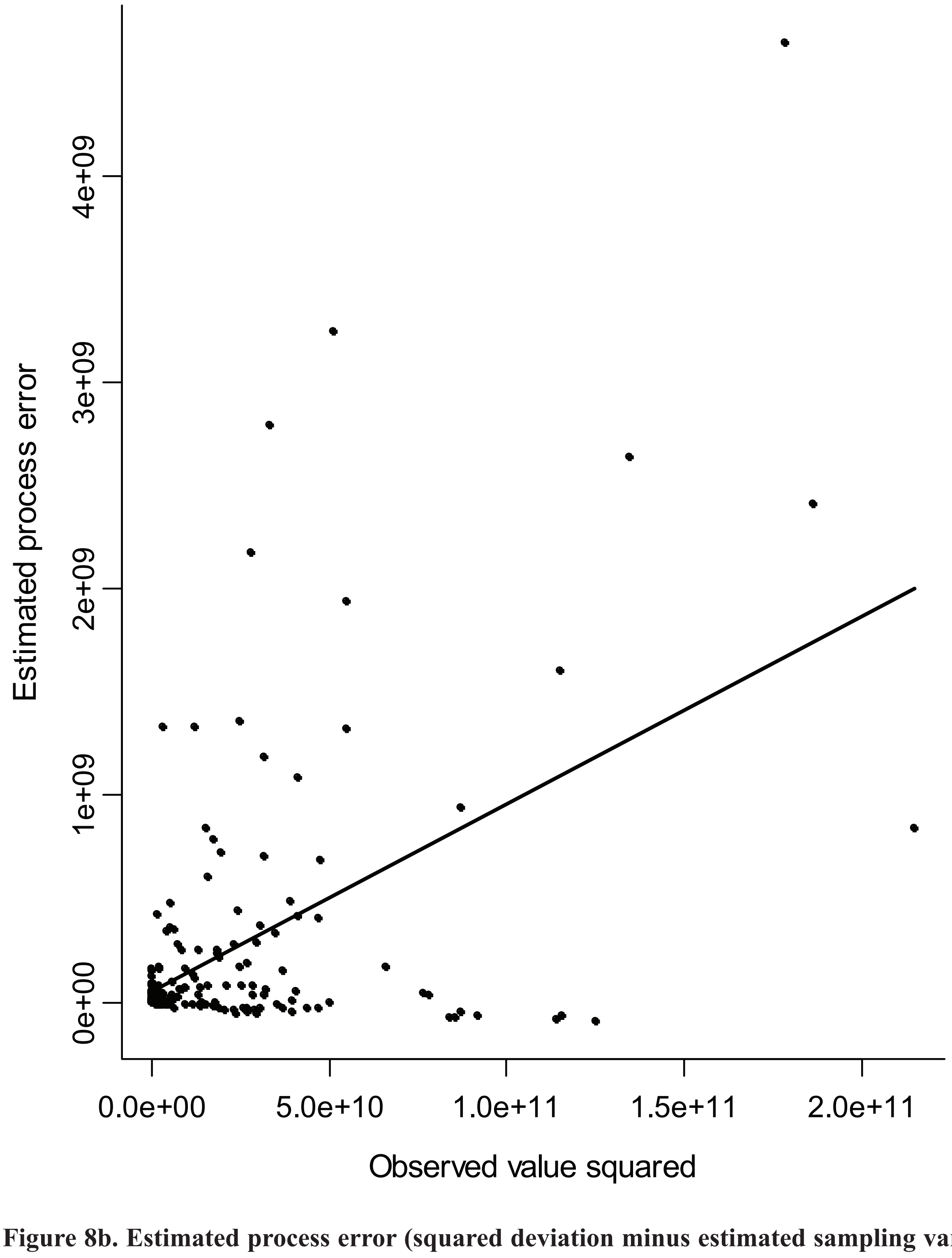

Figure 8b. Estimated process error (squared deviation minus estimated sampling variance) plotted against the square of the observed value of female catch at age.

Appendix A. Selected fishery and survey data summaries.

Table A1. Commercial catch (million pounds, net weight). Figures include IPHC research catches. Sport catch in Areas 2A and 2B is *not* included in this table.

	2A	2B	2C	3A	3B	4	4A	4B	4C	4D	4E	Total
1974	0.52	4.62	5.60	8.19	1.67	0.71	---	---	---	---	---	21.31
1975	0.46	7.13	6.24	10.60	2.56	0.63	---	---	---	---	---	27.62
1976	0.24	7.28	5.53	11.04	2.73	0.72	---	---	---	---	---	27.54
1977	0.21	5.43	3.19	8.64	3.19	1.22	---	---	---	---	---	21.88
1978	0.10	4.61	4.32	10.30	1.32	1.35	---	---	---	---	---	22.00
1979	0.05	4.86	4.53	11.34	0.39	1.37	---	---	---	---	---	22.54
1980	0.02	5.65	3.24	11.97	0.28	0.71	---	---	---	---	---	21.87
1981	0.20	5.66	4.01	14.23	0.45	---	0.49	0.39	0.30	0.01	0.00	25.74
1982	0.21	5.54	3.50	13.52	4.80	---	1.17	0.01	0.24	0.00	0.01	29.01
1983	0.26	5.44	6.38	14.14	7.75	---	2.50	1.34	0.42	0.15	0.01	38.39
1984	0.43	9.05	5.87	19.77	6.69	---	1.05	1.10	0.58	0.39	0.04	44.97
1985	0.49	10.39	9.21	20.84	10.89	---	1.72	1.24	0.62	0.67	0.04	56.10
1986	0.58	11.22	10.61	32.80	8.82	---	3.38	0.26	0.69	1.22	0.04	69.63
1987	0.59	12.25	10.68	31.31	7.76	---	3.69	1.50	0.88	0.70	0.11	69.47
1988	0.49	12.86	11.36	37.86	7.08	---	1.93	1.59	0.71	0.45	0.01	74.34
1989	0.47	10.43	9.53	33.74	7.84	---	1.02	2.65	0.57	0.67	0.01	66.95
1990	0.32	8.57	9.73	28.85	8.69	---	2.50	1.33	0.53	1.00	0.06	61.60
1991	0.36	7.19	8.69	22.93	11.93	---	2.26	1.51	0.68	1.44	0.10	57.08
1992	0.44	7.63	9.82	26.78	8.62	---	2.70	2.32	0.79	0.73	0.07	59.89
1993	0.50	10.63	11.29	22.74	7.86	---	2.56	1.96	0.83	0.84	0.06	59.27
1994	0.37	9.91	10.38	24.84	3.86	---	1.80	2.02	0.72	0.71	0.12	54.73
1995	0.30	9.62	7.77	18.34	3.12	---	1.62	1.68	0.67	0.64	0.13	43.88
1996	0.30	9.54	8.87	19.69	3.66	---	1.70	2.07	0.68	0.71	0.12	47.34
1997	0.41	12.42	9.92	24.63	9.07	---	2.91	3.32	1.12	1.15	0.25	65.20
1998	0.46	13.17	10.20	25.70	11.16	---	3.42	2.90	1.26	1.31	0.19	69.76
1999	0.45	12.70	10.14	25.32	13.84	---	4.37	3.57	1.76	1.89	0.26	74.31
2000	0.48	10.81	8.44	19.27	15.41	---	5.16	4.69	1.74	1.93	0.35	68.29
2001	0.68	10.29	8.40	21.54	16.34	---	5.01	4.47	1.65	1.84	0.48	70.70
2002	0.85	12.07	8.60	23.13	17.31	---	5.09	4.08	1.21	1.75	0.56	74.66
2003	0.82	11.79	8.41	22.75	17.23	---	5.02	3.86	0.89	1.96	0.42	73.19
2004	0.88	12.16	10.23	25.17	15.46	---	3.56	2.72	0.95	1.66	0.31	73.11
2005	0.80	12.33	10.63	26.03	13.17		3.40	1.98	0.53	2.58	0.37	71.82
2006	0.83	12.01	10.49	25.71	10.79	---	3.33	1.59	0.49	2.37	0.37	67.98
2007	0.78	9.74	8.49	26.31	9.42	---	2.81	1.41	0.55	2.72	0.58	62.81

Table A2. Commercial CPUE (net pounds per skate).

Values before 1984 are raw J-hook catch rates, with no hook correction. 1983 is excluded because it consists of a mixture of J- and C-hook data. No value is shown for area/years after 1980 with fewer than 500 skates of reported catch/effort data. Total column recomputed in 2007 with new bottom area numbers.

	2A	2B	2C	3A	3B	4A	4B	4C	4D	4E	Total
J-hook CPUE:											
1974	59	64	57	65	57	---	---	---	---	---	---
1975	59	68	53	66	68	---	---	---	---	---	---
1976	33	53	42	60	65	---	---	---	---	---	---
1977	83	61	45	61	73	---	---	---	---	---	---
1978	39	63	56	78	53	---	---	---	---	---	---
1979	50	48	80	86	37	---	---	---	---	---	---
1980	37	65	79	118	113	---	---	---	---	---	---
1981	33	67	145	142	160	158	99	110	---	---	---
1982	22	68	167	170	217	103	---	91	---	---	---
1983	---	---	---	---	---	---	---	---	---	---	---
C-hook CPUE:											
1984	63	148	314	524	475	366	161	---	197	---	357
1985	62	147	370	537	602	333	234	---	330	---	400
1986	60	120	302	522	515	265	---	427	239	---	356
1987	57	131	260	504	476	341	220	384	---	---	349
1988	134	137	281	503	655	453	224	---	201	---	392
1989	124	134	258	455	590	409	268	331	384	---	376
1990	168	175	269	353	484	434	209	288	381	---	334
1991	158	148	233	319	466	471	329	223	398	---	328
1992	115	171	230	397	440	372	278	249	412	---	336
1993	147	208	256	393	514	463	218	257	851	---	392
1994	93	215	207	353	377	463	198	167	480	---	326
1995	116	219	234	416	476	349	189	---	475	---	351
1996	159	226	238	473	556	515	269	---	---	---	413
1997	226	241	246	458	562	483	275	335	671	---	419
1998	194	232	236	451	611	525	287	287	627	---	425
1999	---	213	199	437	538	500	310	270	535	---	394
2000	263	229	186	443	577	547	318	223	556	---	412
2001	169	226	196	469	431	474	270	203	511	---	379
2002	181	222	244	507	399	402	245	148	503	---	378
2003	184	231	233	487	364	355	196	105	389	---	349
2004	145	212	240	485	328	315	202	120	444	---	340
2005	155	197	203	446	293	301	238	91	379	---	314
2006	147	202	170	403	292	241	218	72	280	---	284
2007	121	172	164	410	261	213	230	66	216	---	268

Table A3. IPHC setline survey CPUE of legal sized fish in weight (net pounds per skate).

Figures refer to entire areas. For cases where only part of an area was fished (e.g., northern 2B, western 3A), the CPUE shown is an adjusted value. *No hook corrections* are applied; J-hook values are raw J-hook catch rates. Area 4EBS is the eastern Bering Sea shelf, first surveyed in 2006. For other years, the 4EBS CPUE is a constructed value based on the NMFS trawl survey and the single 2006 setline data point.

	2A	2B	2C	3A	3B	4A	4B	4C	4D	4EBS	Total
J-hook surveys:											
1974	---	---	---	---	---	---	---	---	---	---	---
1975	---	---	---	---	---	---	---	---	---	---	---
1976	---	---	---	---	---	---	---	---	---	---	---
1977	---	13	---	58	---	---	---	---	---	---	---
1978	---	18	---	27	---	---	---	---	---	---	---
1979	---	NA	---	41	---	---	---	---	---	---	---
1980	---	25	---	76	---	---	---	---	---	---	---
1981	---	16	---	131	---	---	---	---	---	---	---
1982	---	21	114	130	---	---	---	---	---	---	---
1983	---	18	142	119	---	---	---	---	---	---	---
1984	---	25	---	176	---	---	---	---	---	---	---
C-hook surveys:											
1984	---	57	260	361	---	---	---	---	---	7	---
1985	---	42	260	378	---	---	---	---	---	8	---
1986	---	38	283	305	---	---	---	---	---	9	---
1987	---	NA	---	---	---	---	---	---	---	10	---
1988	---	NA	---	---	---	---	---	---	---	20	---
1989	---	NA	---	---	---	---	---	---	---	13	---
1990	---	NA	---	---	---	---	---	---	---	14	---
1991	---	NA	---	---	---	---	---	---	---	12	---
1992	---	NA	---	---	---	---	---	---	---	11	---
1993	---	93	---	261	---	---	---	---	---	22	---
1994	---	NA	---	254	---	---	---	---	---	17	---
1995	29	148	---	300	---	---	---	---	---	20	---
1996	---	156	306	317	352	---	---	---	---	25	---
1997	35	139	411	331	414	237	282	71	111	23	166
1998	---	82	232	281	435	310	216	---	---	30	157
1999	37	88	204	241	438	290	203	---	---	27	147
2000	---	93	233	272	373	282	216	---	215	20	142
2001	41	102	237	256	357	205	171	---	197	21	133
2002	33	92	261	299	297	174	119	---	263	13	128
2003	22	73	223	229	262	158	104	---	195	18	108
2004	27	86	173	270	236	142	73	---	132	18	106
2005	28	72	171	276	211	111	86	---	69	17	99
2006	16	59	144	232	181	88	95	---	63	18	86
2007	19	57	140	212	191	69	87	---	57	13	79

Appendix B. Evolution of IPHC assessment methods, 1982-2007

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 (Quinn et al. 1985). 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. Accumulated data showing very similar trends in catch 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 (Clark and Hare 2003).

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 it 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.

Apart from a few minor and inconsequential corrections and alterations, the 2005 analytical assessment was the same as the 2004 assessment. The only important change in procedure was the use of the NMFS trawl survey to estimate biomass in Area 4CDE where an analytical assessment was not done.

In 2006, growing concerns about migration of legal-sized fish from western to eastern areas led the staff to doubt the validity of the closed-area assessments that had been done for many years (Clark and Hare 2007a). The staff therefore estimated coastwide abundance by fitting the model to a coastwide dataset, and estimated biomass in each area in accordance with survey estimates of relative abundance (Clark and Hare 2007b). The 2007 assessment followed the same procedure. Sublegal discard mortality in the halibut fishery was added to the removals included in the assessment; it had the effect of decreasing the present biomass estimate by less than 1%.

References

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., 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.. 2003. Assessment of the Pacific halibut stock at the end of 2002. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2002:95-120.

Clark, W. G., and Hare, S. R.. 2004. Assessment of the Pacific halibut stock at the end of 2003. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2003:171-200.

Clark, W. G., and Hare, S. R. 2007a. Motivation and plan for a coastwide assessment. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:83-96.

Clark, W. G., and Hare, S. R. 2007b. Assessment of the Pacific halibut stock at the end of 2006. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:97-128.

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.

Quinn, T. J. II, Deriso, R. B., and Hoag, S. H. 1985. Methods of assessment of Pacific halibut. Int. Pacific Halibut Comm. Sci. Rep. 72.

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.

Effect of station depth distribution on survey CPUE

William G. Clark

Abstract

IPHC setline surveys stations are set on a 10 nmi grid in depths from 20 to 275 fm, so the depth distribution of the survey stations should approximate the depth distribution of the bottom in each regulatory area, and the simple mean CPUE should be the same as the depth-stratified mean CPUE. This is true in all areas except Area 2A, where the depth-stratified mean is consistently higher than the simple mean, by an amount averaging 40%.

Background

In the 2006 assessment (Clark and Hare 2007) the staff estimated coastwide abundance by fitting the standard assessment model to a coastwide data set, and then estimated exploitable biomass in each regulatory area by apportioning the total in proportion to an estimate of stock distribution derived from the setline survey. Specifically, an index of abundance in each area was calculated by multiplying setline CPUE (running 3-year average) by total bottom area between 0 and 300 fm. The logic of this index is that survey CPUE can be regarded as an index of density, so multiplying it by bottom area gives a quantity proportional to total abundance. The estimated proportion in each area is then the index value for that area divided by the sum of the index values.

The survey CPUE value used for each area in each year is the simple mean of the CPUE values recorded at all stations fished in that area. The stations are set on a 10 nmi grid between 20 and 275 fm so they provide a uniform coverage of every area. As a result, the various conditions that could affect CPUE—such as depth, substrate type, temperature, dissolved oxygen—should be represented in the stations in approximately the same proportions as in the area as a whole. The simple mean of all the station values should therefore be close to the mean that would be obtained by stratifying the stations by e.g. depth, computing a mean CPUE for each depth stratum, and then computing an overall CPUE by weighting the stratum means by the actual proportion of bottom in each depth stratum.

At the 2007 annual meeting there was some concern expressed that the simple mean CPUE failed to account for the variation in CPUE with depth. This paper addresses that concern. The depth distribution of survey stations is compared with the depth distribution of the bottom in each area, and the depth-stratified mean CPUE is compared with the simple mean CPUE. They are the same in all areas except Area 2A, where the depth-stratified mean is 40% higher.

Depth distribution of survey stations, bottom, and commercial catch

Figure 1 shows the cumulative depth distribution of survey stations and bottom in each regulatory area. As expected they are quite similar but there are some differences. There is a small excess of shallow stations in Area 2A, a small excess of deep stations in Area 2C, and a substantial excess of shallow stations in Area 4B. The comparison is not shown for Area 4C because it is not surveyed, nor for Area 4D which is partially surveyed.

For information, Figure 2 shows the depth distribution of commercial catch in each area as a check on whether the maximum survey depth of 275 fm is deep enough. It is. More than 98% of the commercial catch is taken at depths less than 280 fm in all areas except Area 4A, where the proportion is 94%.

Variation of survey CPUE with depth

The effect of any mismatch between survey station depth distribution and bottom depth distribution in a given area will depend on how survey CPUE varies with depth in that area. The effect of depth was estimated by fitting a generalized additive model to survey CPUE from 2001-2006 in which year entered as factor to account for changes over time and depth entered as a smooth function. Set time and soak time were also included in initial fits but neither was significant so they were dropped. The smooth functions estimated for the depth effect are plotted in Figure 3.

In most areas the variation with depth while highly significant is not very large. The exceptions are Area 2A and to a lesser extent Area 2C. In Area 2A catch rates are low in shallow water, so the excess of shallow stations in Area 2A can be expected to have an important effect.

Unstratified and depth-stratified survey CPUE series

Figure 4 shows the bottom line: a comparison of simple mean CPUE series and depth-stratified CPUE series for each area. The two are very close in all areas except Area 2A, where the depth-stratified CPUE is consistently higher, by an amount averaging 40%.

Discussion

Except for Area 2A, the results presented here show that there is no need to calculate a depth-stratified survey CPUE. Because of the uniform station distribution, the simple mean CPUE accounts for the depth effect in all areas except Area 2A. This is a good outcome, because computing a depth-stratified estimate would complicate the assessment and possibly raise some statistical problems. In Area 2A some adjustment may be appropriate, but it is difficult to choose a number. The Area 2A survey CPUE has a high sampling variance either way it is computed. The depth-stratified CPUE exceeds the simple mean by an average of 40%, but that excess has a standard deviation of 10%, so an approximate 95% confidence interval for the adjustment is 20-60%.

References

Clark, W. G., and Hare, S. R. 2007. Assessment of the Pacific halibut stock at the end of 2006. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:97-128.

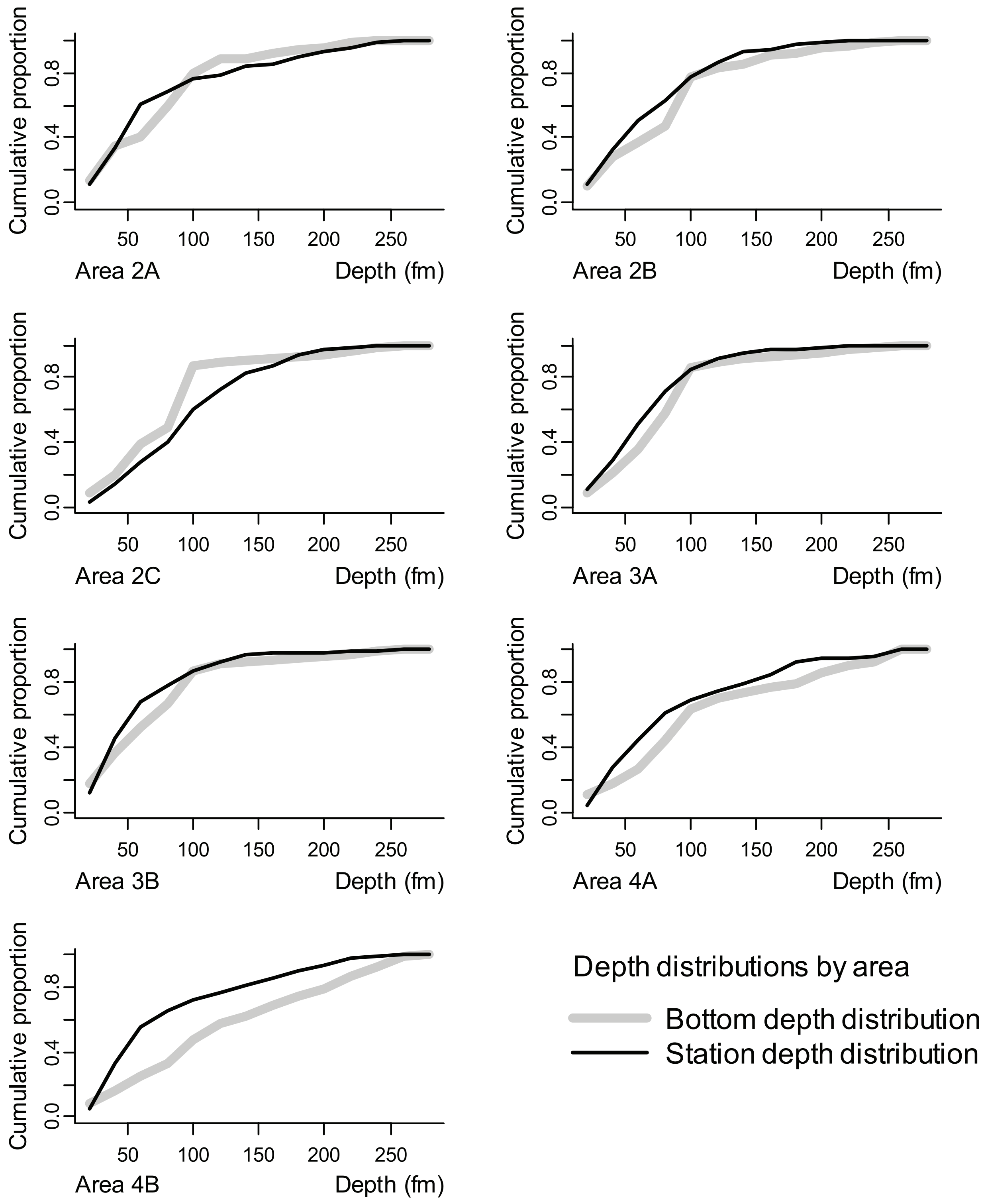

Figure 1. Cumulative distribution of bottom depth and survey station depth in each area.

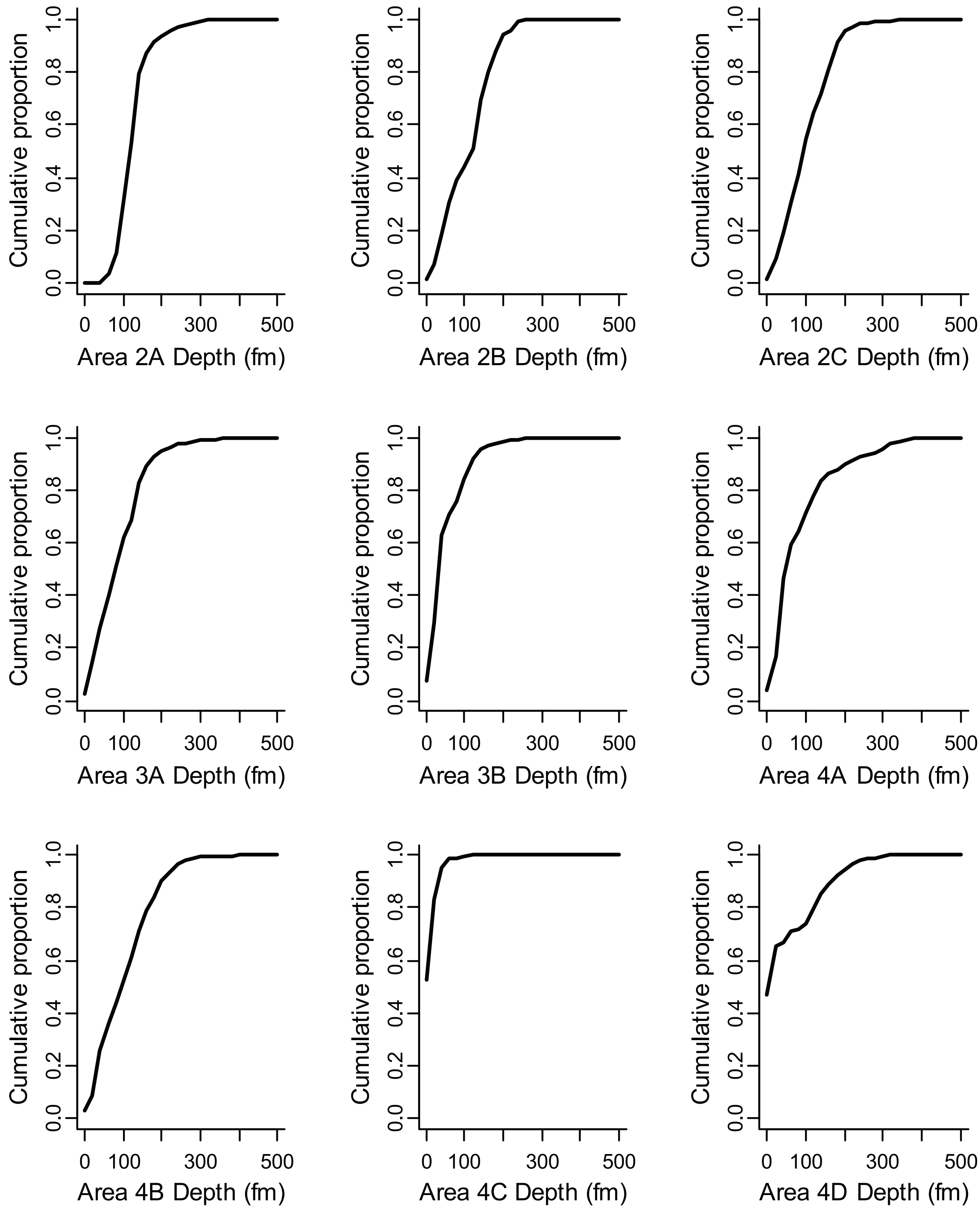

Figure 2. Cumulative depth distribution of commercial catch (2004-2006) in each area.

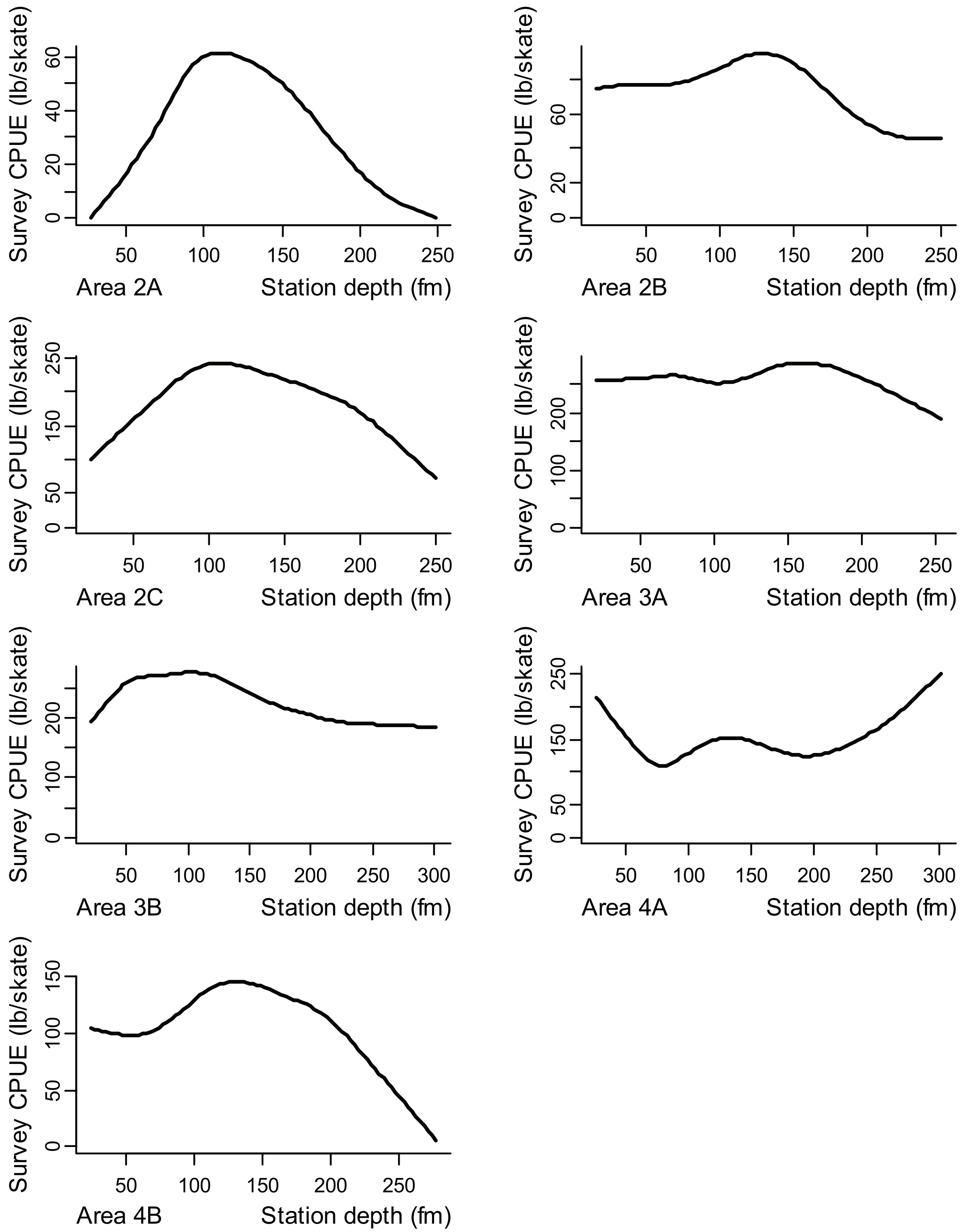

Figure 3. Variation of survey CPUE (2001-2006) with depth in each area. The plotted line is the depth term from a fitted generalized additive model with year as a factor and depth as a smooth function.

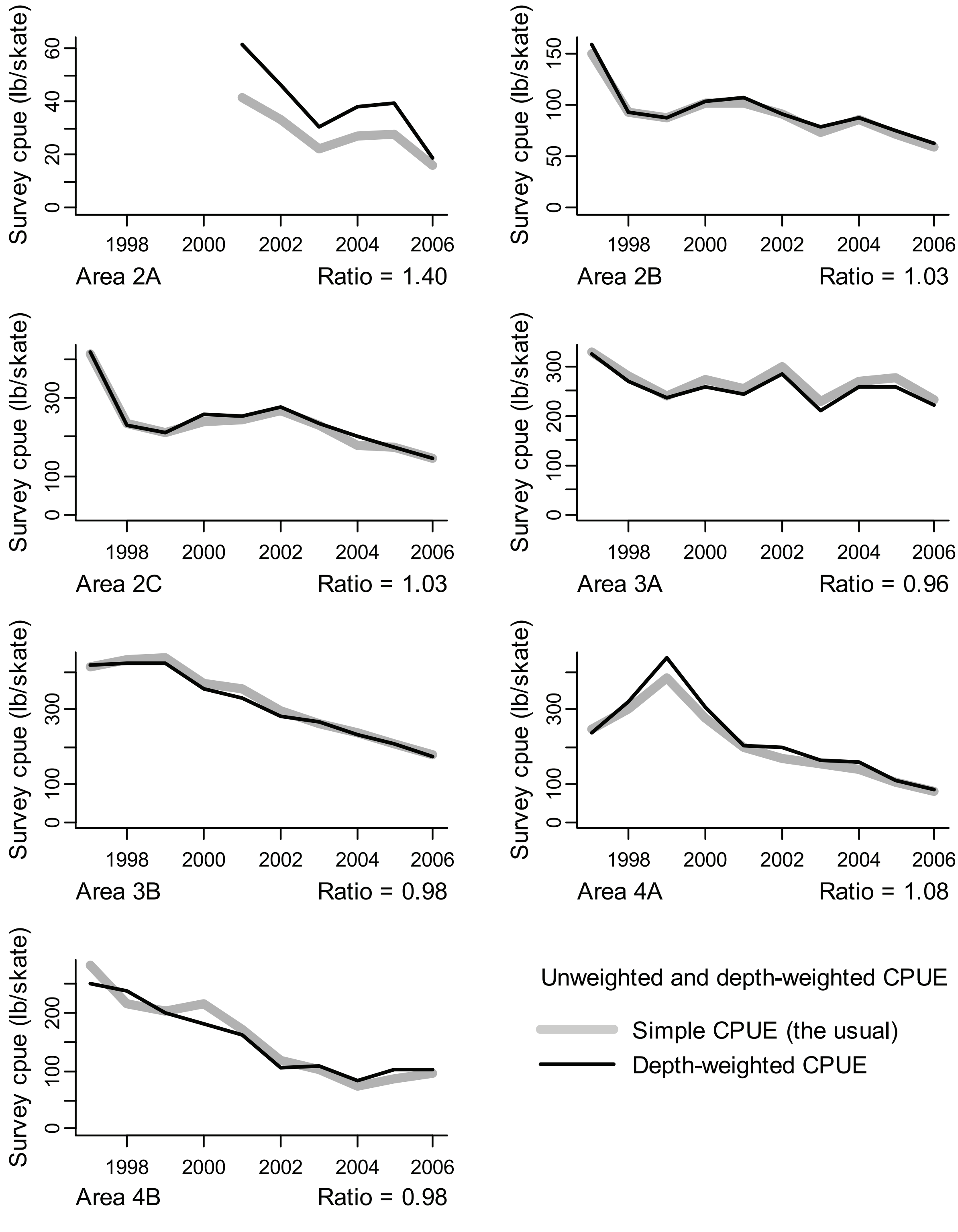

Figure 4. Comparison of simple and depth-stratified survey CPUE series in each area. The ratio shown at the lower right of each graph is the sum of the stratified estimates divide by the sum of the simple means.

Effect of hook competition on survey CPUE

William G. Clark

Abstract

The catch of halibut at setline survey stations is reduced by other species that take the bait. Hook competition is highest in Area 2A and lowest in Areas 4B and 4D. Despite some differences in catch composition, overall hook competition varies little among Areas 2B, 2C, 3A, 3B, and 4A.

Background

In the 2006 assessment (Clark and Hare 2007) the staff estimated coastwide abundance by fitting the standard assessment model to a coastwide data set, and then estimated exploitable biomass in each regulatory area by apportioning the total in proportion to an estimate of stock distribution derived from the setline survey. Specifically, an index of abundance in each area was calculated by multiplying setline CPUE (running 3-year average) by total bottom area between 0 and 300 fm. The logic of this index is that survey CPUE can be regarded as an index of density, so multiplying it by bottom area gives a quantity proportional to total abundance. The estimated proportion in each area is then the index value for that area divided by the sum of the index values.

This procedure assumes that survey catchability (the constant of proportionality between CPUE and the density of halibut on the bottom) is the same in all areas. But the CPUE of halibut is reduced by the number of baits taken by other species, and if the strength of this effect varies among areas the result would be differences in survey catchability among areas. In particular, it has long been suspected that the large number of dogfish caught in Area 2B depressed survey CPUE of halibut there, and almost certainly it does. Similarly, the large number of cod caught in Area 4 can be presumed to lower the CPUE of halibut there.

This paper reports estimates of the effect of hook competition on the survey CPUE of halibut in each regulatory area. It turns out that the effect is very similar in all regulatory areas except 2A and 4B, so the simple procedure used in the 2006 assessment was sound for the most part.

Mathematical treatment of hook competition

The sequence of events that occurs when a baited longline is set and various species go after the bait has been studied theoretically and experimentally for decades (Sigler 2000 and references therein). Mathematically the process of baits being removed from a longline by different species is the same as the process of fish being removed from a population by different fisheries and natural predators. We can represent each kind of bait taker as removing a certain proportion of the baits per unit time, so that the number of baits B_i taken by a given species *i* during a soak time *T* is given by the familiar catch equation:

$$B_i = F_i \cdot B_0 \cdot \left(1 - \exp\left(-Z \cdot T\right)\right) / Z$$

where F_i is the instantaneous rate of bait removal by species i, B_0 is the initial number of baited hooks, and $Z = \sum_j F_j$ is the sum of the instantaneous rates applied by all bait takers. This formulation has been found to describe quite well the actual sequence of catches during the first few hours of soaking in experiments where the time of each capture was recorded by a hook timer (Sigler 2000; Somerton and Kikkawa 1994). After the first few hours the rates of bait removal by all takers drops off, either because the remaining bait has lost its scent or because the bait takers in the vicinity of the gear have been depleted. This has also been observed for halibut in unpublished hook timer experiments conducted by IPHC (Steven Kaimmer, IPHC, pers. comm.). Beyond a certain point, therefore, soak time does not matter.

In the IPHC setline survey every string soaks for at least five hours, and there is no significant difference in CPUE between shorter and longer soaks. (Set time is also insignificant.) Soak time can therefore be regarded as effectively the same for all survey sets, and the term T can be left out of the bait removal equation.

The instantaneous rate of bait removal by halibut can be taken to be proportional to the local density of halibut, and depending on size and gear selectivity some proportion of halibut that take a bait will also be hooked and caught, so the catch per skate of halibut C_h will be proportional to the density of halibut D_h multiplied by the last term in the bait removal equation:

$$C_h = k \cdot B_h = k \cdot F_h \cdot B_0 \cdot \left(1-\exp\left(-Z\right)\right)/Z = k' \cdot D_h \cdot B_0 \cdot \left(1-\exp\left(-Z\right)\right)/Z$$

where k and k' are constants of proportionality. In this equation, $\left(1-\exp\left(-Z\right)\right)$ is the fraction of baits removed by all takers during the active period, and $\left(1-\exp\left(-Z\right)\right)/Z$ is the average number of baits remaining over the course of the active period as a proportion of the initial number. If this term is the same in all areas, then survey CPUE is a consistent index of density across areas. Otherwise survey CPUE does not index density consistently across areas. Equivalently, if the fraction of baits taken is the same in all areas, then survey CPUE is a consistent index of density.

It is interesting to note that the effect of hook competition on the comparability of survey CPUE is wholly determined by the total bait removal rate Z. The species composition of the bait takers makes no difference. If 80% of the baits are taken in both Area X and Area Y (meaning that Z is the same), and the catch in Area X is all halibut and the catch in Area Y is half halibut and half dogfish, the survey CPUE's of halibut in the two areas will accurately reflect the relative densities of halibut.

Comparison of bait removal rates among areas

Figure 1 shows raw hook count data from the setline survey by area. In most areas 10-20% of the bait is recovered. The exceptions are Area 2A, where only 7% is recovered, and Area 4B, where a third of the bait is recovered. That is also true of stations on the Bering Sea edge in Areas 4A and 4D, but the overall rates for Area 4A and 4CDE are in the usual range. The recovery rate for all areas combined is 14%, which is a little low because in recent years all hooks have been counted in Area 2B while only 20% have been counted elsewhere, and the Area 2B recovery rate is 12%.

Table 1 shows the bait removal rates calculated from the raw data in Figure 1. Halibut are minor players in all areas except 2C, 3A, and 3B. The "Other" category, dominated by whatever species send back empty hooks or skins, are the major players in all areas. The last column is the ratio of the coastwide to the area-specific value of $\left(1-\exp\left(-Z\right)\right)/Z$. It is the multiplier that should be applied to each area's survey CPUE to make all of them consistent.

Variances of the correction factors were calculated by the jackknife method, leaving out one year at a time, on the grounds that year-to-year changes are the major sources of variance in survey data. The standard deviations of the correction factors were 0.05 for Area 2A, 0.02 for Areas 2B, 2C, 3A, and 3B, and 0.01 for Areas 4A, 4B, and 4CDE. So as a statistical matter they are almost all significantly different from one, but in most cases there is very little practical difference.

Discussion

While the standard catch equation performs reasonably well in predicting the timing of catches during the first few hours of soaking, it is not perfect. The rate at which baits are taken has been found to vary over time in different ways for different gears and target species, and Sigler (2000) reports substantial variation among stations. The estimate of the average number of baited hooks fishing during the active period is based on the observed proportion of bait recovered and an assumed rate of decline based on the catch equation. It is at therefore an approximation, and small differences among areas in the estimate are not reliable even though they may be statistically significant.

Hook competition is not the only possible cause of differences among areas in setline survey catchability. Differences in temperature may affect survey catchability directly or indirectly (Stoner et al. 2006). The availability of natural prey may also affect the desirability of the bait (Stoner 2004). Factors such as these are almost surely responsible for the substantial year-to-year variability of survey CPUE and may also cause some consistent differences among areas.

References

Clark, W. G., and Hare, S. R. 2007. Assessment of the Pacific halibut stock at the end of 2006. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2006:97-128.

Sigler, M. F. 2000. Abundance estimation and capture of sablefish (*Anoplopoma fimbria*) by longline gear. Can. J. Fish. Aquat. Sci. 57:1270-1283.

Somerton, D. A., and Kikkawa, B. S. 1994. A stock survey technique using the time to capture individual fish on longline gear. Can. J. Fish. Aquat. Sci. 52:260-267.

Stoner, A. W. 2004. Effects of environmental variables on fish feeding ecology: implications for the performance of baited fishing gear and stock assessment. J. Fish. Biol. 65:1445-1471.

Stoner, A. W., Ottmar, M. L., and Hurst, T. P. 2006. Temperature affects activity and feeding motivation in Pacific halibut: implications for bait-dependent fishing. Fish. Res. 81:202-209.

Table 1. Bait recovery fractions, total instantaneous bait removal rates *Z*, and various species-specific removal rates F_i, by area. The last column shows the multiplier that should be applied to each area's survey CPUE to make survey CPUE a consistent index of density across areas. Except as noted below, all data from 2001-2006 were used. The "Other" category in this table includes emtpy hooks and skins.

				Instantaneous rates by species					
Area	Stations	Fraction recovered	Z	Halibut	Cod	Dogfish	Sablefish	Other	Correction factor
2A	504	0.07	2.66	0.06	0.00	0.09	0.21	2.30	1.25
2B	1014	0.12	2.16	0.14	0.00	0.27	0.09	1.65	1.07
2C	653	0.15	1.89	0.28	0.01	0.08	0.12	1.40	0.97
3A	2222	0.10	2.30	0.47	0.07	0.38	0.10	1.29	1.12
3B	1328	0.15	1.92	0.55	0.18	0.01	0.07	1.12	0.98
4A	664	0.18	1.70	0.23	0.24	0.00	0.03	1.20	0.91
4B	528	0.34	1.09	0.08	0.10	0.00	0.01	0.90	0.72
4D	336	0.34	1.07	0.10	0.20	0.00	0.00	0.77	0.71
4CDE*	204	0.18	1.74	0.06	0.29	0.00	0.00	1.39	0.92
EBS*	82	0.13	2.03	0.04	0.29	0.00	0.00	1.69	1.02
All	7313	0.14	1.97	0.25	0.07	0.18	0.08	1.38	1.00

* The Area 4CDE data are all stations fished in the eastern Bering Sea in 2006, including the shelf stations, island stations, 4D edge, and 4A edge. The "EBS" data are just the eastern Bering Sea shelf stations fished in 2006.

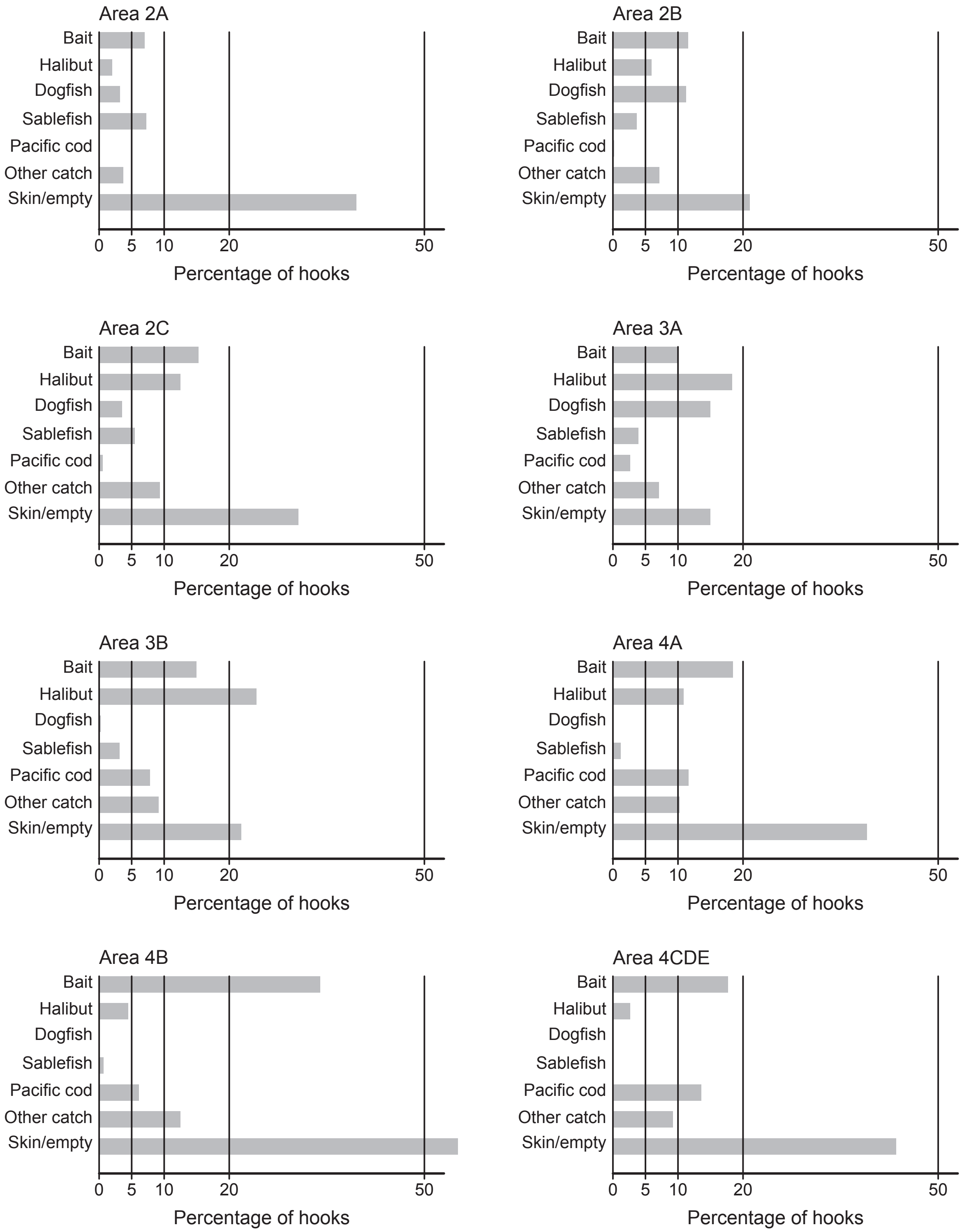

Figure 1. Survey hook contents by area, 2001-2006 data combined.

Effect of migration on achievement of proportional harvest under a system of survey apportionment of total catch

William G. Clark

Abstract

Apportioning catches among areas in proportion to a snapshot of biomass distribution will not necessarily result in equal fishing mortality rates in all areas if there is net migration occurring. But the differences are at most very small at the rates of migration likely to occur in halibut, and if the snapshot of biomass distribution is taken at midseason, when the IPHC setline survey is done, there is no difference. The survey apportionment of halibut biomass and yield should result in equal rates of instantaneous fishing mortality in all areas even in the presence of migration.

Background

In the 2006 stock assessment, owing to concerns about the effect of migration on closed-area assessments, the staff estimated coastwide biomass with a coastwide model fit and then apportioned the coastwide estimate among regulatory areas in proportion to a survey index of biomass in each area (survey CPUE multiplied by bottom area out to 300 fm). This apportionment relies strongly on the assumption of equal survey catchability among areas, which is reasonable but questionable. And even if survey catchability is equal, there is some question as to whether the survey apportionment of yield will really achieve the Commission's goal of proportional harvest among areas when fish are migrating before and after the survey. This paper addresses that question.

The first step is to define terms precisely. In fishery science the terms "exploitation rate" and "harvest rate" have the same technical meaning, which is the catch as a fraction of the initial number present or $u = C / N_0$ where u is the rate of exploitation, N_0 is the number of fish present in an area at the beginning of the season, and C is the catch in number during the season (Ricker 1975). When fishing and natural mortality are treated as continuous processes, and there is no net migration, the rate of exploitation is determined by the Baranov equation, which is $u = F \cdot \left(1 - e^{-(F+M)}\right) \big/ \left(F + M\right)$ with F denoting the instantaneous rate of fishing mortality and M the instantaneous rate of natural mortality. The instantaneous rate of total mortality is $Z = F + M$ and the number of fish surviving to time t during the season is $N_t = N_0 \cdot e^{-Z \cdot t}$. With Z substituted for $\left(F + M\right)$, the Baranov catch equation is $C = u \cdot N_0 = N_0 \cdot F \cdot \left(1 - e^{-Z}\right) \big/ Z$.

Migration complicates the calculations but not greatly. The effect of emigration on the number of fish present in an area is the same as the effect of an addition to natural mortality, and the effect of immigration is the same as the effect of a reduction in natural mortality. Let X denote the instantaneous rate of net emigration from an area. Then $Z = F + M + X$. The Baranov equation still holds, but a higher rate of fishing mortality F is needed to achieve a given exploitation rate u in competition with emigration as well as natural mortality. Likewise if there is net immigration into an area at a rate X, we have $Z = F + M - X$ and a lower F can achieve the same exploitation rate.

The Commission's goal of proportional harvest really means achieving the same rate of instantaneous fishing mortality in all areas. It would not be proper to achieve the same rate of exploitation in all areas if that required applying a high rate of instantaneous fishing mortality in emigrant areas and a low rate in immigrant areas. Instead the aim is for fish to encounter the same rate of instantaneous fishing mortality wherever they go.

From this point of view the yield apportionment proposed by the staff is suspect, because it applies the same exploitation rate to both emigrant and immigrant areas without any consideration of the implied instantaneous rates of fishing mortality. The remainder of this paper reports calculations of the implied instantaneous rates.

Apportionment based on stock distribution at the beginning of the season

Consider as a simple hypothetical example a stock that occupies two regulatory areas of equal size, Area 1 and Area 2. Suppose that an equal number $N1 = N2 = 1000$ recruits appears in each area at the beginning of the year, and that we have good estimates of the equal numbers in the two areas at that time. Further suppose that our target rate of exploitation is $u^* = 0.20$, and we set a target catch of 200 fish from each area accordingly. If there is no net migration between the areas, and if natural mortality $M = 0.15$, the instantaneous rate of fishing mortality in both areas will be $F^* = 0.242$, the intended rate.

But suppose there is a net emigration from Area 1 to Area 2 at a rate $X = 0.05$, and as a simplifying approximation treat this as immigration into Area 2 at the same rate $X = 0.05$. The rates will be the same at the start of the year, but by the end of the year emigration at a 0.05 rate from Area 1 will be slightly less than a 0.05 rate of immigration into Area 2 (about 0.045). The calculations will therefore slightly overstate the effects of migration, but the approximation is quite close and it greatly simplifies the analysis.

The number of fish in Area 1 as the year progresses will be governed by $Z1 = F1 + M + X$. The value of $F1$ required to achieve the target catch there is $F1 = 0.248$. In Area 2 the number of fish is governed by $Z2 = F2 + M - X$, and the required fishing mortality is $F2 = 0.235$. If X were 0.10 rather than 0.05, the resulting rates would be $F1 = 0.255$ and $F2 = 0.229$. So there is some disparity in realized instantaneous fishing mortality rates in this case, but it is small.

Apportionment based on stock distribution during the season

The apportionment used by the staff is based on estimated relative abundance at the time of the survey, which is halfway through the year. At that time the number of fish present is $N_0 \cdot e^{-Z/2}$. This midyear number is naturally very close to the average number present during the course of the year, which is $N_0 \cdot \left(1 - e^{-Z}\right) / Z$, so the effect of an apportionment based on midyear abundance can be studied using either expression.

In our hypothetical case, midyear abundance in Area 1 will be lower than in Area 2 because $Z1 > Z2$, so less than half the yield will be allocated to Area 1. This will move both $F1$ and $F2$ toward the target rate F^*. The actual rates that will result from this apportionment must satisfy two conditions: the catch from each area must be proportional to the value of $N_0 \cdot \left(1 - e^{-Z}\right) / Z$ in each area, and the sum of the catches must equal the target catch from both areas. Written out, the second condition is:

$$F1 \cdot N1 \cdot \left(1 - e^{-Z1}\right) \Big/ Z1 \ + \ F2 \cdot N2 \cdot \left(1 - e^{-Z2}\right) \Big/ Z2 \ = \ F* \cdot \left(N1 + N2\right) \cdot \left(1 - e^{-Z*}\right) \Big/ Z*$$

It is clear from this equation that $F1$ and $F2$ must be equal, because otherwise the catches would not be proportional to the values of $N_0 \cdot \left(1 - e^{-Z}\right) \big/ Z$. And in this case they must both be nearly equal to $F*$ in order to achieve the target catch because $\left(1 - e^{-Z}\right) \big/ Z$ is nearly linear in Z and $Z*$ is midway between $Z1$ and $Z2$. Simulation results bear this out; the equilibrium values of $F1$ and $F2$ are equal, and their value is equal to $F*$ to three decimal places. So a midyear apportionment results in fishing at the same instantaneous rate in both areas even when there is some net migration.

The results from this simple example extend to the case of several areas with possibly different sizes and possibly unbalanced migration rates because an equation like the one above must be satisfied by the operation of the midyear apportionment procedure. Using subscript a to denote area, the equation is:

$$\sum_a F_a \cdot N_a \cdot \left(1 - e^{-Z_a}\right) \Big/ Z_a = F* \cdot \left(\sum_a N_a\right) \cdot \left(1 - e^{-Z*}\right) \Big/ Z*$$

All of the $\{F_a\}$ must be equal; denote this value $\tilde{F}$. The equation can then be written:

$$\tilde{F} \cdot \sum_a N_a \cdot \left(1 - e^{-Z_a}\right) \Big/ Z_a = F* \cdot \sum_a N_a \cdot \left(1 - e^{-Z*}\right) \Big/ Z*$$

The summation on the left is just the coastwide sum of the average numbers in all areas. The summation on the right is the sum of the average numbers that would occur if all of the $\{Z_a\}$ were equal to $Z*$. At any moderate levels of migration these two values will be nearly equal, so $\tilde{F}$ will be nearly equal to $F*$.

References

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191: 382 p.