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. Their reports, submitted in July, are posted in the stock assessment section of the IPHC website: http://www.iphc.washington.edu/halcom/research/sa/sa.html. 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 likelihood-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. Report of the June 2007 stock assessment workshop. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2007 (this volume).

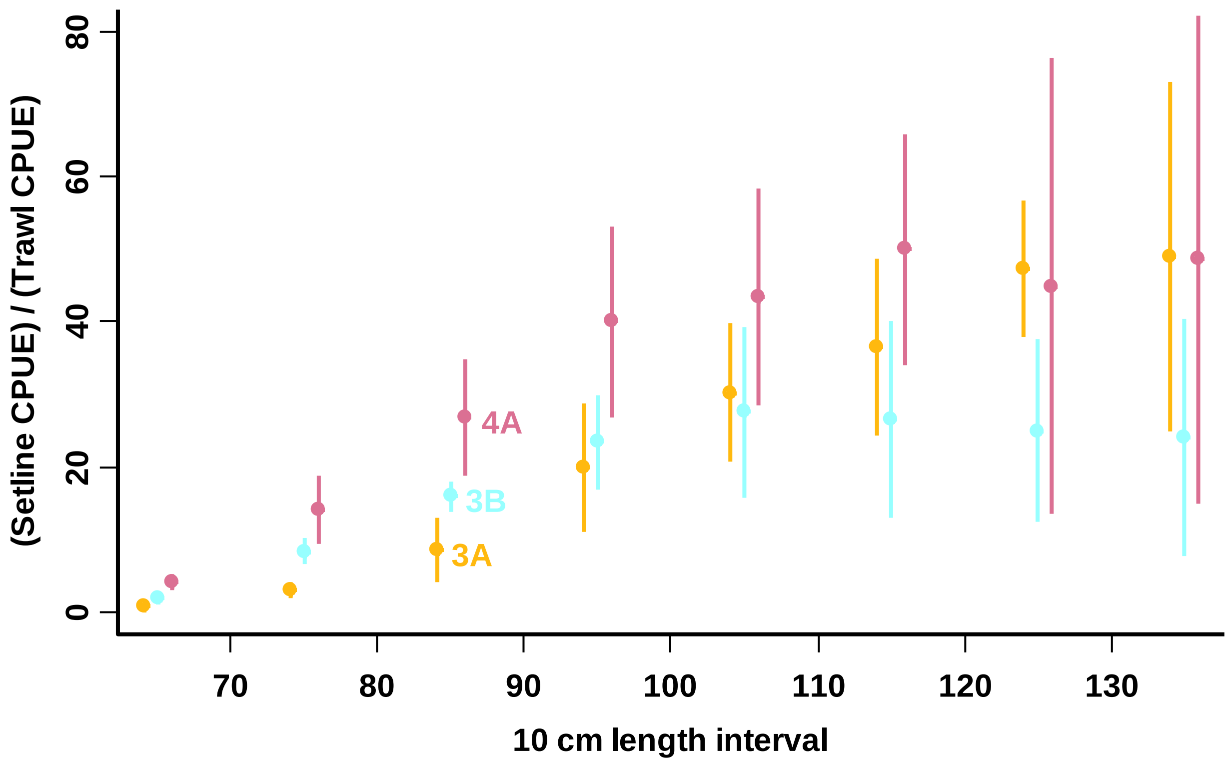

Figure 1. Ratios of setline to trawl survey CPUE at length by area.