Sources of uncertainty in annual CEY estimates

William G. Clark, Steven R. Hare, and Bruce M. Leaman

Abstract

This paper summarizes the development of IPHC catch limits and discusses the major elements of uncertainty in them, with special reference to the reliability of estimates of present abundance relative to the historical minimum.

Introduction

The present IPHC harvest policy is to limit total annual removals to 20% of exploitable biomass, a target called the Constant Exploitation Yield or CEY. Under an alternative policy put forward by the staff, called a Conditional Constant Catch or CCC policy, removals would be limited to the lesser of a target constant catch or 25% of exploitable biomass. Both policies are designed to achieve a large proportion of maximum yield in the long term while assuring that spawning biomass will remain above the historical minimum reached in the 1970s (when strong year-classes were produced despite relatively low spawning biomass).

The exploitable biomass estimates and the harvest policy are both based on the stock assessment. Along with an estimate of present biomass, the assessment produces estimates of historical abundance and recruitment. These historical estimates are the basis of the stock and fishery simulations that are conducted to evaluate alternative harvest policies (Fig. 1).

There are a number of uncertainties in both the stock assessment and the simulations, and they have different effects on different quantities of interest. For example, the working value of natural mortality has a large effect on the estimate of present biomass in absolute terms, but no effect on the estimate of present biomass relative to the historical minimum, because its effect is to scale all the abundance estimates similarly.

This paper attempts to catalogue (briefly) all sources of uncertainty in the annual estimates of CEY (or CCC) and their effect on: (1) estimates of present abundance, (2) estimates of present abundance relative to the historical minimum, and (3) estimates of stock productivity (i.e., the average long-term yield available at different constant harvest rates, including maximum sustainable yield MSY). Of these, the second is clearly the critical one because the Commission's paramount conservation goal is to maintain a healthy spawning stock. The following sections discuss the various sources of uncertainty; they are summarized in Table 1.

Statistical variability

Fitting the stock assessment model to the commercial and survey data consists of locating the best estimates of a number of parameters, the important ones being commercial catchability and selectivity, survey catchability and selectivity, and the initial abundance of each year-class present in the data series. Like all other statistical estimates, these estimates of model parameters have some variance due to random variability in the data (e.g., in age composition estimates and survey CPUE). But this random estimation error in the halibut assessment is only a few percent—negligible—because we are fitting models with a modest number of parameters (called "parsimonious" models) to a large number of data points. *Given a particular model,* therefore, the parameter estimates are very well determined and their statistical variance is very small.

Like the parameter estimates themselves, the variance estimates are conditional on the particular model being fitted. If the model is wrong, the estimated variances will be too low whether they are computed with the usual asymptotic approximation or by bootstrapping (Punt and Butterworth 1993). Still, the internal variances of the halibut parameter estimates are so low that statistical variability can be ignored even if the variances are underestimated by a factor of two.

Despite the small variances, a model can produce quite different estimates of, say, recent year-class strengths when another year of data is added to the fit, and different plausible models can do the same thing when fitted to the same data series. The year-to-year and model-to-model variability result from a different kind of uncertainty—model specification uncertainty, discussed below mainly in relation to survey catch rates.

Natural mortality

Historical estimates of the abundance of a year-class can be thought of as back-calculations of abundance at each age based on estimates of total subsequent fishery removals and natural deaths (e.g., total number alive at age 10 = total deaths at age 10 and older). The working value of the natural mortality rate therefore has a large effect on historical abundance estimates and through them on present abundance estimates.

Natural mortality is virtually impossible to estimate, so the working value (presently 15%) is almost certainly wrong and so are the absolute abundance estimates, but it turns out that the consequences are not serious. As mentioned above, the estimate of present biomass relative to the historical minimum is reliable. Estimates of stock productivity are not much affected by the natural mortality rate because the natural deaths are added in during back-calculations of stock abundance in the assessment and then subtracted out during the forward calculations of stock productivity in the fishery simulations. Estimates of long-term yield at a given harvest rate, including MSY, are therefore very robust to uncertainty in the natural mortality rate, at least for Pacific halibut (Clark 1999).

But what if natural mortality is not a single rate but different rates according to age, sex, and year? What if there was a large change in natural mortality coincident with the large change in recruitment after the 1977 regime shift, or coincident with the large decrease in growth rates in recent years? Differences of that sort cannot be detected or estimated. The number of possibilities is enormous, so it would not be a simple matter to formulate a set of plausible alternatives for routine consideration. The effect of various patterns could (and should) be investigated by conducting simulations as described in the discussion section below. Our estimates of abundance and productivity are probably not robust to large temporal swings in natural mortality.

Survey catchability and selectivity

The most important differences among IPHC assessment models in recent years have related to survey catchability and selectivity. Present abundance estimates are determined mostly by survey catch rates because we assume that there has been a constant proportional relationship since 1974 between average survey catch rates and abundance. (Abundance in 1974 is determined entirely by the catch at age data and the natural mortality rate, so in effect the survey catch rates are calibrated by the first part of the data series and then used to infer abundance in the latter part.) The devil is in the details of the proportional relationship.

When survey data were first used in the assessment in 1995, and for a few years thereafter, there were two models of survey catch rates. In what was called the "length-specific" version, vulnerability to capture was assumed to depend on length and a length-specific selectivity function was estimated that increased from zero at around 60 cm to one (100%) at 130 cm or less. In what was called the "age-specific" version, vulnerability to capture was assumed to depend on age and an age-specific selectivity function was estimated that increased from zero at around age 5 to one at age 17 or less. The catchability coefficient in these (and all other) models is the coefficient of proportionality between the abundance of fully selected fish and the survey catch rate of those fish (in our case, numbers of fish per standard skate). It has the same value in both models, so the only difference was in the assumed determinant of selectivity: length or age.

Various measures of goodness of fit provided no reason to prefer one model over the other (Clark and Parma 1999) so both fits were reported, but the CEY estimates and catch limit recommendations were based on the age-specific version because it produced lower estimates of present biomass. Poor retrospective performance of the age-specific model in recent years and independent data from the NMFS trawl survey (Clark 2003) have now shown that the length-specific model is in fact much more credible, and it will be used for the 2004 CEY estimates.

Another instance of competing models of survey data occurred in the 1999 assessment, when it was suspected that a change in survey bait had increased survey catchability in 1993. The effect was uncertain at the time of the assessment but again the CEY estimates were based on a model that assumed the suspected bait effect and therefore produced the lower abundance estimates. (An experiment done in 2000 showed that the effect of the bait change was negligible, and the adjustment was removed from the assessment.)

Similarly in the 2002 assessment when several models were fitted to account for anomalies that appeared in the 3A assessment, the lowest estimates of present abundance were used to calculate CEY (Clark and Hare 2003).

At present the question of survey selectivity appears to be settled in favor of a length-based schedule, but catchability is always a question. Is the survey catchability of a large halibut really the same as it was twenty years ago? Has the increase in dogfish in Area 3A, for example, reduced the catchability of halibut? Has the increase in halibut abundance reduced feeding opportunities and made a baited longline more attractive, thereby increasing catchability? Does bait quality have a large influence on a given year's survey CPUE? The generally good agreement between trends in setline and trawl survey CPUE of large fish gives some reason to believe that setline survey catchability has not changed dramatically over the last twenty years, but both survey data series are quite noisy and there is no assurance that catchability has not changed or will not change in the future. Any appreciable change in catchability would derail the assessment, because more than anything else the assessment relies on the survey CPUE of fully selected fish to index relative abundance.

Commercial catchability and selectivity

Before 1995 the assessment was done with CAGEAN, which did not use survey data at all but relied on commercial CPUE in the same way that the present assessment relies on survey CPUE. In particular it assumed constant catchability and age-specific selectivity, and when age-specific commercial selectivity changed owing to decreased growth in the early 1990s, CAGEAN began to perform poorly.

In the present model commercial catchability and age-specific selectivity parameters are allowed to change over time, so real changes can be accommodated in the model as they occur in the data series. The rate of change is limited, however, so the commercial data act to damp variations in the abundance estimates resulting from year-to-year swings in the noisier survey data. But it is the trend of the survey data that determines the estimated long-term change in relative abundance and therefore the estimate of present absolute abundance.

Other elements of model specification

This year the staff will attempt to fit a sex-specific model, and that will raise a number of questions about possible differences between females and males in natural mortality, catchability, and selectivity. These differences, if they exist, will probably not be detectable or at least not distinguishable owing to lack of data on the sex composition of the commercial catches and the strong interaction between natural mortality and selectivity estimates. But such differences cannot be ruled out, and their operation would presumably bias the estimates of a model that assumes no differences.

Recruitment and growth dynamics

Our present working hypotheses about stock dynamics are that recruitment is largely determined by the environment (over the range of observed stock sizes) and that growth is density-dependent (Clark and Hare 2002). These hypotheses do not enter the stock assessment in any way so they do not affect our estimates of abundance, but they are the whole basis of the fishery simulations and harvest policy. If they are wrong, the harvest policy will not perform as expected. In particular, if productivity is much lower than we believe, the harvest policy could eventually result in driving the stock down to a level near its historical minimum, which would require a drastic curtailment of the fishery. The stock would not be put at risk because the assessment would monitor its decline, but the fishery would. As a practical matter, the stock is so far above that level now that it would take some time to get there, and before that could happen there would be a change in our working hypotheses and a course correction in our harvest policy.

Ecosystem change

Apart from our specific working hypotheses about stock dynamics, we make the basic assumption that past stock performance informs us about present and future performance. Is that really true? How will global warming affect halibut in Canada and Alaska in the decades ahead? How has the development of other groundfish fisheries affected halibut, or the decline of Steller sea lions? If history is not a reliable guide, what is an appropriate harvest policy? In particular, what are appropriate reference points for minimum spawning biomass and maximum fishing mortality?

Discussion

The major source of uncertainty in the annual CEY estimates is model specification, meaning how the model goes about predicting the observations, especially the survey data. Once the model is specified, the parameter and abundance estimates are well determined by the data. Over the last ten years there have been a number of occasions when the model specification was in doubt, and the staff has dealt with the uncertainty by basing its recommendations on the model that produced the lowest estimates. As a routine procedure, this method of dealing with uncertainty can be expected to cost the fishery some yield from time to time, but it serves the stock well.

The problem of competing plausible models is very common in fisheries. Some authors (e.g. Punt and Hilborn 1997) advocate attempting to assign a probability or relative weight to every plausible model and then computing some kind of probability distribution for the estimate of abundance. We had no basis for choosing such weights on any of the occasions when the halibut model was in doubt, so we believe this approach would have been arbitrary, controversial, and risky in our case.

Like all other assessment models, ours is a highly simplified representation of one part of a complex and dynamic ecosystem (Schnute and Richards 2001). Natural mortality probably does differ by sex and vary with age and time, but we do not know and cannot estimate the differences, so we use a single value in the model. Survey catchability probably has varied over the years, but if we allowed it to vary in the model it would be confounded with the recruitment estimates and we would not get useful estimates of either, so we treat it as constant, in effect estimating an average value. We are obliged to keep the assessment model simple (parsimonious) so that we can estimate the parameters.

It is reasonable to think that even though it is a simplification, the fitted model does an adequate job of approximating stock trends and present relative abundance, but that is not certain. The fact is that any model that can be fitted will be misspecified in some respects, and we do not know how that will affect our abundance estimates, harvest policy evaluations, and eventually the stock and fishery. The question can be studied by building much more detailed and complex models (called operating models) to generate test data and then fitting the parsimonious assessment model to the data to generate estimates of historical and present abundance. A harvest policy can then be chosen based on the historical estimates and applied to the estimates of present abundance to set catch limits, and those catches can be taken from the stock in the operating models year after year to see how well the assessment and harvest policy perform when the data come from a system that does not match the assessment model.

This approach has been used successfully elsewhere (Cooke 1999, Punt and Smith 1999). It consists of two parts: the operating model (really a suite of operating models), and a management procedure, which consists of all the steps between the data produced by the operating model and the recommended catch limit. The management procedure can be more or less elaborate. It may be a simple rule based on CPUE (like IPHC quota setting in the old days), or a fitted production model, or a delay-difference model (Parma 2002), or a conventional age-structured assessment *cum* harvest policy evaluation like ours. The aim is to test different management procedures and find one that works well with a variety of operating models and is therefore robust to uncertainty about the detailed workings of the real stock.

References

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

Clark, W. G. 2003. Estimates of juvenile halibut abundance and distribution in Area 3 from NMFS trawl surveys. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2002.

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

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

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

Parma, A. M. 2002. In search of robust harvest rules for Pacific halibut in the face of uncertain assessments and decadal changes in productivity. Bull. Mar. Sci. 70:423-453.

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

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

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

Schnute, J.T., and L.J. Richards. 2001. Use and abuse of fishery models. Can. J. Fish. Aquat. Sci. 58:10-17.

Table 1. Effects of various sources of uncertainty on estimates of stock abundance and productivity.

Source of uncertainty	**Effect on estimates of present biomass**	**Effect on estimates of present biomass relative to minimum**	**Effect on estimates of stock productivity (yield curve)**
Statistical variability of parameter estimates	Slight	Slight	Slight
Natural mortality rate	Major	Slight	Slight
Assumptions about survey catchability and selectivity	Major	Major	Slight
Assumptions about commercial catchability and selectivity	Minor	Minor	Slight
Other elements of assessment model specification	Major	Major	Slight
Working hypotheses about recruitment and growth dynamics	None	None	Major
Ecosystem change	None	None	Major

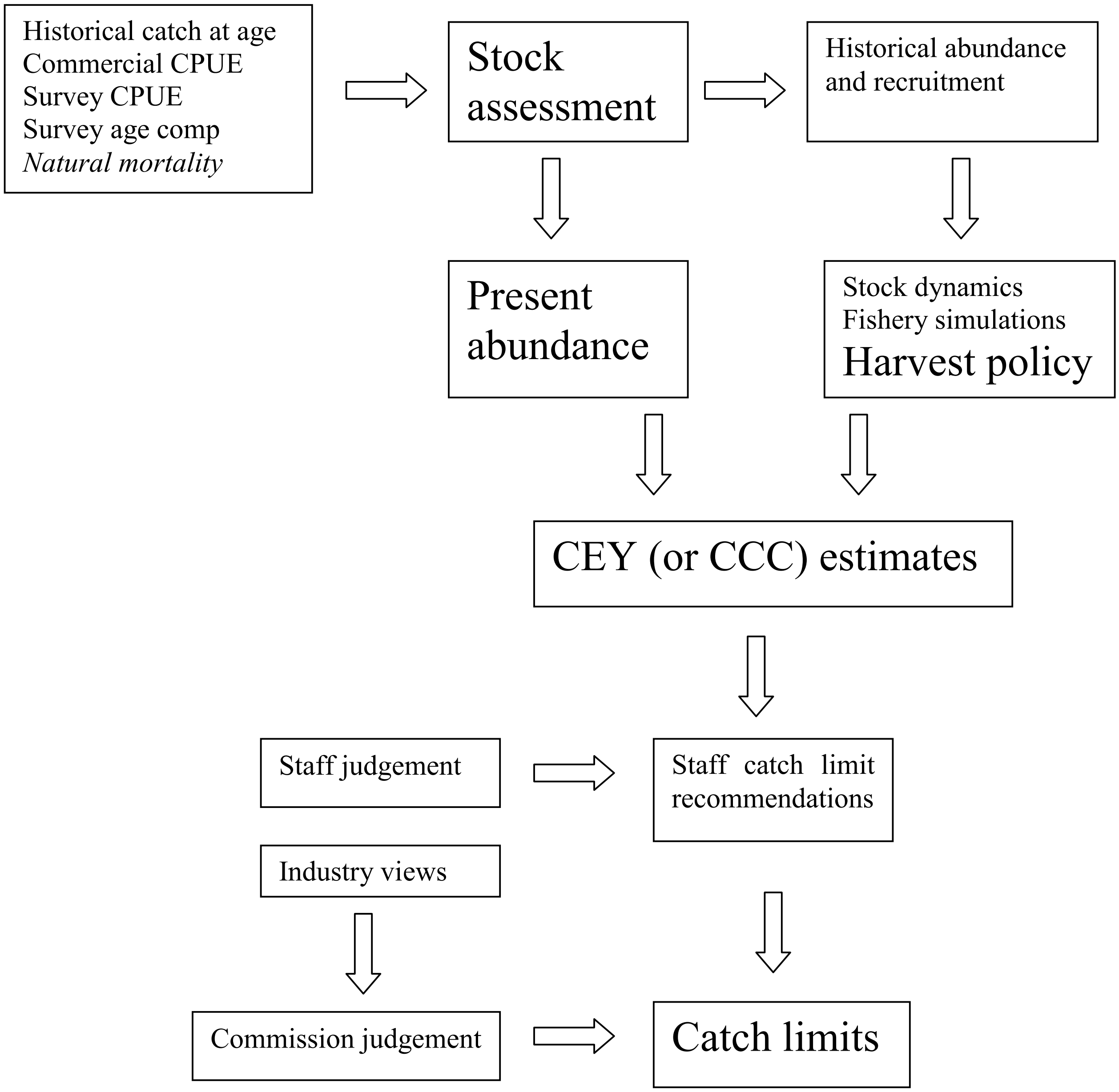

Figure 1. Evolution of IPHC catch limits.