UM Independent System for Peer Reviews
Consultant Report on:

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

June 27-28, 2007

Paul A. Medley
Sunny View
Jack Hole
Alne, YO61 1RT
UK
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:

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

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

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

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.

$$A = 2\pi r^2 - 2r^2 \,\mathrm{Arcsin}(b/r) - 2ab$$

Where r = radius of the effective area fished, a = 0.5 times the distance between hooks and $b = (r^2\text{-}a^2)^{0.5}$. 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.

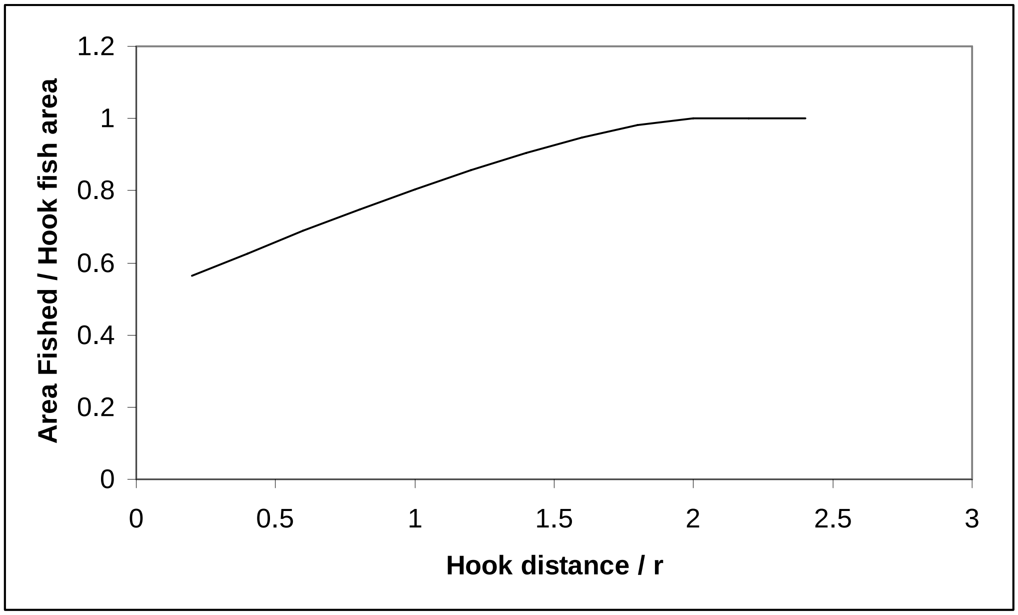

Figure 1 The asymptote is reached when the hook distance is twice the radius of the effective area fished. If this pattern is observed, it strongly suggests that catchable fish are distributed evenly and move relatively slowly.

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?