Report on the 2007 Assessments of Aleutian Islands Atka Mackerel and Pollock

NIWA Client Report: WLG2008-050
June 2008

NIWA Project: NTV08301

Report on the 2007 Assessments of Aleutian Islands Atka Mackerel and Pollock

Author
R.I.C.C. Francis

Prepared for

The Center for Independent Experts

NIWA Client Report: WLG2008-050
June 2008

NIWA Project: NTV08301

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

Approved for release by:

Dr Andrew Laing

Executive Summary

A panel reviewed the 2007 assessments of Aleutian Islands Atka mackerel and pollock. The Panel met 9-12 June 2008 at the Alaska Fisheries Science Center of NOAA/NMFS in Seattle. The initial draft assessments, together with related material, were presented to the Panel, additional analyses were requested and carried out, and the Panel discussed the results.

The analysts' understanding and use of the available data were generally good and the two assessments were basically sound and based on science that was close to the best available.

Some suggestions are made that might incrementally improve future assessments. The main generic topics covered related to survey catchability, data weighting, estimating selectivities, ageing error, additions to the assessment software (AMAK), and documentation. Other topics addressed for the individual assessments included the delineation of the assessment area, model selection, spatial structure, and the use of tagging data.

1. Background

This report reviews, at the request of the Northern Taiga Ventures, Inc. (see Appendix 1), the 2007 assessments of Aleutian Islands Atka mackerel and pollock. The author was provided with various documents (Appendix 2) and participated in the meeting which considered the assessment, and in subsequent discussions with the assessment team.

2. Review Activities

The review panel met 9-12 June 2008 at the Alaska Fisheries Science Center of NOAA/NMFS in Seattle. Those attending the meeting included three reviewers, the scientists primarily responsible for the stock assessments, and those presenting related material (Appendix 3). The assessments, and related material, were presented to the Panel, some additional analyses were requested and carried out, and the Panel discussed the results with the assessment team (see Appendix 4 for the agenda).

3. Findings

This was, for me, a most enjoyable review. I found the AFSC staff to be unfailingly helpful and cheerful, and I appreciated the informality of the meeting, which I felt was highly conducive to a successful review. Where there were disagreements, discussions were always amicable, stimulating, and constructive.

I thought that the analysts showed a good understanding of their data and produced sound assessments. Although I have some concerns about some aspects of the assessments, and the way they were presented, none of these substantially affects the main assessment results.

I will first discuss some generic issues, and then turn to the individual assessments.

3.1 Issues common to both assessments

3.1.1 Survey catchability

In both assessments survey catchability appeared to be poorly determined, so it was deemed necessary to make strong, and conservative, assumptions about this parameter. This decision seemed appropriate to me. However, I have two concerns about it.

My first concern relates to the communication of uncertainty. Both assessments were, because of the catchability assumptions, very uncertain, and it seems to me important that this uncertainty should be clearly communicated to users of these assessments. This did not seem to have been done. However, I acknowledge that different fisheries administrations have different conventions about where and how such uncertainty is expressed. What I don't know is where the target audience for these documents expects to see expressions of the extent of uncertainty about individual assessments. For myself, I was surprised to see no mention of this uncertainty (or even catchability) in either the summary document (Plan Team 2007) or the Executive Summary sections of the assessment report. In both reports the assessments are described (in sections on ABC Recommendations/Considerations) as 'conservative' because of the catchability assumptions, and perhaps this is meant to imply uncertainty.

My second concern is that I was not sure that the issue of catchability had been fully explored for these stocks, although this was hard to determine since most of the exploration had been done in previous years. What I wanted to see for each stock was a clear hypothesis about why survey catchability was so poorly determined. My own hypothesis, after a preliminary reading of the assessment reports, was simple. We can estimate survey catchability well only when the survey biomass index shows a clear response to the catches (either a decrease in the index when catches are high, or an increase when they are low). When there is no such obvious response (as was the case for both stocks) we can estimate only an approximate upper bound for the catchability (i.e., the value above which we could expect to see a clear biomass response to the catches). I will discuss below, in the sections relating to the specific stock assessments, why this simple 'no-response' hypothesis turned out to be inadequate for both stocks.

3.1.2 Data weighting

It is important that the appropriate weight is given to each data set in a stock assessment model. This weight is determined primarily by the 'spread' parameter of the assumed error distribution (e.g., the sample size for a multinomial error, and the standard deviation (s.d.) or coefficient of variation (c.v.) for a normal or lognormal error). For at-age data there is also a need to ensure that outlier observations can't receive undue weight, and this is usually achieved by modifying the likelihood in a procedure called 'robustification'.

An important point to note is that the spread parameters must allow for both observation and process error. That is, the total error, whose expected size is specified by the spread parameter, and which may be written as (observed value – model value), is the sum of the observation error (observed value – true value) and the process error (true value – model value), or, more briefly, $\varepsilon_{\text{total}} = \varepsilon_{\text{observation}} + \varepsilon_{\text{process}}$. This is important because we can often estimate a spread for $\varepsilon_{\text{observation}}$ from the data themselves (e.g., the c.v.s in table 15.4 of Lowe et al (2004) and the sample sizes at the top of p. 146 of Barbeaux et al (2007) are for observation errors). However, we must allow for additional spread due to process error. For trawl survey biomass estimates, I suggest using a process-error c.v. of 0.2 (following Francis et al 2003). (Note that c.v.s add as squares, so adding this to an observation-error c.v. of 0.25, say, produces an overall c.v. of $(0.2^2 + 0.25^2)^{0.5} = 0.32$). For multinomial at-age data, the total spread parameters are usually called effective sample sizes, and I have some suggestions for their calculation below (see Section 3.1.5). An advantage of using the lognormal error distribution for at-age data, as I prefer to, is that a process-error c.v. can be estimated as a model parameter.

3.1.3 Estimating selectivities

The strategy used in estimating fishery selectivities in both assessments (and survey selectivities in the pollock assessment) was to estimate many parameters (one for each combination of year and age) but then apply smoothing constraints (in both the age and time domains) to reduce the effective number of parameters and increase the plausibility of the resultant selectivities.

I don't think this is a good approach because of two problems associated with the smoothing constraints. The first is that it's hard to determine how smooth is smooth enough. With no smoothing at all, it is obvious that we would be fitting to noise, and thus biasing our estimates of year-class strengths and estimating far too many parameters. With strong smoothing we are in danger of missing the signal in the at-age data concerning changes in selectivity. There seems to be no objective way of choosing the best the degree of smoothing, so this choice becomes purely subjective and arbitrary. The second problem was described during the review by Ana Parma: smoothing constraints that work in point estimates does not always work in Bayesian estimates from MCMCs (Monte Carlo Markov Chains). That is, the same smoothing parameters that produce 'plausible' point-estimate selectivities may result in implausibly jagged selectivities in MCMS.

My suggestion is to use the above approach, with smoothers, to generate hypotheses about (a) the best parametric form for selectivities for this fishery and (b) how to divide the period of the fishery into blocks of years, within which the selectivity doesn't vary. Then test these hypotheses using tools like the Akaike Information Criterion (Akaike 1974) on models with no smoothers (so selectivity curves are parametric and do not change within year blocks). For example, the pollock assessment results suggest that only two year blocks would be necessary for the fishery selectivity (fig. 1A.23, Barbeaux et al. 2007). This approach will probably not produce quite as good a fit to the at-age data, but I think it will reduce the probability of over-fitting and will allow the use of standard techniques for model selection.

3.1.4 Ageing error

I was surprised to see no allowance in the assessments for ageing error. I grant that making such an allowance is unlikely to have a profound effect on either assessment. However, it should make a small improvement, and I am not aware of any reason not to do this.

The usual approach is to provide (directly or indirectly) an age-misclassification matrix, M, where M_{ij} is the probability that a fish in the jth age class will be assigned to the ith age class because of ageing error. Then the vector, $\hat{P}_{ay}$, of expected proportions at age in year y is replaced in the at-age likelihood by $P'_{ay} = M \text{ x } \hat{P}_{ay}$, where 'x' denotes matrix multiplication. We were shown evidence of imprecision in the ageing of these two species (about 84% agreement for Atka mackerel, and 64% for pollock). I would expect that allowing for this would have a noticeable effect (an increase in variability) on the estimated year-class strengths, especially for pollock. Ideally, the age-misclassification matrix should be constructed using data from a replicate ageing experiment in which each reader follows exactly the same procedure that is used in production ageing for the species. The above percent agreement figures relate to a slightly different situation (in which, I believe, an initial reading from a less-experienced reader was compared to a 'test' reading, for quality-control reasons).

Usually, age-misclassification matrices are used to represent only imprecision, but they can also incorporate bias. The radiometric validation of pollock ageing by Kastelle & Kimura (2006) can be used to estimate bias in annulus-count ages (by fitting a line through the open circles in fig. 1 of this paper) and it would be most interesting to see whether the stock assessment model fitted better with or without this bias. [I note that the Kastelle & Kimura study can only really document relative bias between annulus-count and radiometric ages, and that it is a matter of debate as to which method is more likely to be biased. The authors seem to suggest that the fault is with the radiometric ages, but I can see no reason to rule out bias in the annulus counts.]

3.1.5 AMAK modifications

AMAK is already a very useful and quite flexible stock-assessment program. It's main lack at the moment is thorough documentation (see Section 3.1.6). I'd like to suggest that consideration be given to some other modifications, in addition to an allowance for ageing error, as described above (see Section 3.1.4).

AMAK outputs two diagnostics that are useful in judging whether the data weighting was appropriate (in the sense discussed in Section 3.1.2): 'Average Effective *N*' for at-age data sets, and 'RMSE' for biomass indices (e.g., see table 15.8 in Lowe et al. 2007). I suggest changes to both of these. We discovered, during the review, that the method used to calculate the former was not the best, and I suggested two alternative methods. I have now extended those methods to deal with the situation in which the multinomial sample sizes vary with year (see Appendix 5). What I suggest outputting is the correction factor f, where $N_{\mathrm{corr},y} = fN_{\mathrm{init},y}$. The advantage of outputting f, as opposed to the current mean($N_{\mathrm{corr},y}$), is that it is more easily interpretable. I'll explain what I mean by that after discussing my second suggestion, which is to replace the RMSE statistic with what in New Zealand is called the standard deviation of the normalised residuals (SDNR). The idea is to 'normalise' each residual, so that it's expected distribution is that of the standard normal distribution, and then calculate a standard deviation (e.g., with lognormal errors the standardisation involves taking logs and then dividing by the s.d. – in log space – of the error distribution). The interpretations of f and SDNR are similar. Both should be close to 1 if the data weightings are correct. A value much larger (or smaller) than 1 indicates that the assumed error distribution has too little (or too much) spread. (To increase/decrease the spread of an error distribution we should decrease/increase the sample size, if the error is multinomial, and increase/decrease the c.v., if it is lognormal or normal.) For reasons discussed in Appendix 5, it may be desirable to output two versions of f.

For me, a primary characteristic of a good stock-assessment program is flexibility, by which I mean that the user should be offered a wide range of options. Some obvious options to add to AMAK are sex-specific observations and more choice of error distributions (e.g., lognormal, Fournier, and Coleraine errors for at-age data, and lognormal – as opposed to normal-log – for biomass indices; see Bull et al (2008) for definitions of these distributions). Incidentally, since the standard formula for calculating a survey biomass estimate assumes that the expected value of that estimate is proportional to the true biomass, I think the normal-log likelihood used in AMAK for these observations should be replaced by the true lognormal likelihood.

I find profiling to be a very powerful tool for understanding what is driving a stock assessment, and how different data sets, and parameters, interact. It would be good to simplify, and make more robust, the calculation of profiles from AMAK. I'm not sure how profiling is done currently, but it appeared to be time-consuming, and was not very robust (some parts of the profiles presented during the review seemed not to have converged). The user should easily be able to plot the relationships between the profiled parameter and (a) all likelihood components, and (b) any other estimated parameter.

Finally, I suggest trying to make the objective function in AMAK more clearly Bayesian, and less ad hoc. This would involve removing most, if not all, of the weighting parameters (denoted λ_k in table A-3 of Lowe et al. 2007) and allowing the provision of priors for all estimated parameters, and not just a few (users may default to the uniform prior in many cases, but they should not be forced to). Note that, from a Bayesian point of view, the term L_3 in table A-3 of Lowe et al. 2007 should really be expressed as a prior on the recruitment parameters, and the strength of the prior should be defined by the s.d. σ_R, instead of the weight λ_3. Allowing the provision of priors for all parameters raises interesting issues as to default priors and the choice of parameterisation. The default in AMAK seems to be the uniform prior, and this seems to be interpreted as no prior. To see that this is wrong, note that assuming a uniform prior in estimating q is not the same as choosing to estimate log(q), instead of q, with a uniform prior. I think the latter is better.

3.1.6 Documentation

I felt that the reviewers' task was significantly hindered by poor documentation, both of the stock assessments and AMAK, the software used. Considerable time was spent during the review clarifying issues that should have been well described in the documents provided to the reviewers for preliminary reading. Some very minor issues were still not explained by the end of the meeting.

Assessment documentation

One of the reviewers' tasks is to evaluate and comment on the appropriateness of the various decisions (including assumptions) made by the analyst in assessing the stock. Thus there is a need for these decisions, and the reasons for them, to be clearly specified in the assessment reports. I did not find this to be so for these assessments. I am aware of the time constraints involved in doing these assessments. However, since both assessments differed only in relatively small ways from those for previous years, it should have been possible to provide better documentation.

Some things that were not clear in one or both of the assessments were: the weightings applied (in the form of multinomial sample sizes, coefficients of variation (c.v.s), and likelihood weights, λ_k) to each set of observations; the rationale for restricting some data sets to subareas; exactly which parameters were estimated for each model (by this I mean primary parameters, not derived parameters, like $F_{35\%}$, – see example in Table 1); which model results (including c.v.s and confidence intervals) were from point estimates, and which from MCMCs (and if from MCMCs, whether estimates were means, medians, or modes of posterior distributions); and the reasons for preferring one model over another. Further details are given for the individual assessments below.

Table 1: My attempt to describe the (primary) parameters estimated in model 4 of the Atka mackerel assessment. Note that I found only 389 parameters, but table 15.8 of Lowe et al. (2007) says there were 410.

Type	Symbol(s)	Number	Comments
Fishing mortality	$\varphi_i\ \mu^f$	31	For years 1977-2007
Survey catchability	q^s (or μ^s?)	1	
Survey selectivity	s_j^s (or η_j^s?)	10	For ages 1–10
Fishery selectivity	s_j^f (or η_{ij}^f?)	300	30 years (1977-2006) x 10 ages
Mean recruitment	R_0	1	
Recruitment variability	σ_R	1	
Recruitment deviates	ε_i	45	For years 1963-2007
		389	

Another good reason for good documentation is that the discipline required in producing it helps to clarify issues in the mind of the analyst. During the review meeting it sometimes required an extended discussion to elicit from the analysts the reasons for one of their decisions.

One difficulty for the reviewers was that some of the decisions most important to these assessments were made in previous years, and the rationale for these decisions (where provided), and the analyses on which they were based, were provided only in previous assessment reports. The reviewers' task, and the discussions in the review meeting, would have been much easier had the basis for all major decisions been fully described in a single document, perhaps one prepared specifically for the review.

AMAK documentation

I commend the AFSC for having the foresight to provide AMAK as a standard stock assessment program for use in many assessments. Such standard programs have many benefits, including greater assurance that major bugs have been detected and eliminated, simplification of the training of assessment staff, and greater ease in comparing different assessments and transferring skills developed with one species to another. However, the value of such standard software is much reduced when, as is the case with AMAK, it is not comprehensively documented. This lack of documentation made it more difficult for analysts to describe their model assumptions, and for the present reviewers to understand what they were reviewing. Examples of some (but not all) features for which I could find no documentation were the method of smoothing year-to-year variation in selectivities and the calculation of effective sample sizes. What was perhaps worse is that some of the equations provided were inaccurate (e.g., those for proportions at age in the fishery and the multinomial likelihood – both of which fail to correspond to what occurs in the AMAK code). I note that the only AMAK documentation I was able to get before the review (and then only by asking for it) was dated 2004, and thus, presumably, not current.

Ideally, any assessment model should be fully documented by the associated input files and the software manual. The former list all the data used, together with the assumptions made by the analyst, and the latter provides a means of interpreting the files and describing the consequences of all assumptions.

3.2 Atka mackerel assessment

3.2.1 Survey catchability

In this assessment, survey catchability, q^s, was estimated to be 1.5, which I found troublesome and puzzling. It's troublesome for two reasons. Once we allow the possibility that $q^s > 1$ (which is certainly conceivable, but not, I believe, very likely) then it's difficult to know whether any particular value is conservative (in the sense that $q^s = 1$ *must* be conservative when are sure that $q^s \leq 1$). Also, the estimate appeared to be strongly dependent on what seemed to be an entirely arbitrary prior distribution. Increasing (or decreasing) the standard deviation of the prior would directly increase (or decrease) the estimate of q^s.

What I found puzzling was that, contrary to what would be predicted by the simple 'no-response' hypothesis I described above (see Section 3.1.1), the estimate of q^s was greater than 1. I think this needs more exploration, and the formulation of a new hypothesis. As a starting point I suggested a profile on q^s, which was constructed during the review, but which I felt we didn't fully explore (there were also some convergence problems with the profile). My notes say that all data sets supported $q^s > 1$ in this profile, but this seems not to have been true in earlier assessments: in 2004 it was true only of the fishery at-age data (cf Models 3 and 4 in table 15.9 of Lowe et al (2004)). (There should be more information about this in 2003 assessment report, but this was not provided to the reviewers.)

3.2.2 Assessment area

I was confused about what area was associated with the stock being assessed. The title of the assessment report (Lowe et al. 2007) suggested that it was just the Aleutian Islands (AI); table 15.1 of that report made it clear that East Bering Sea (EBS) catches were included (and this was reinforced by the label 'BSAI' on the ABCs and TACs in this table); but I inferred from table 15.8 that BS (Bering Sea) stations were excluded from the survey biomass (and, apparently, also from the survey at-age data).

From a discussion during the review I concluded that BSAI was better than AI as an area descriptor for this stock. Given that, it seems to me illogical to exclude the survey data from the BS stations. This is particularly relevant because it effects the consequences of assumptions about survey catchability. I understood that the reason for excluding this area was that the associated biomass estimates are too uncertain, often depending strongly on the presence or absence of a single large catch. Such uncertainty is certainly annoying, and inconvenient, but it does not seem a good reason to exclude the area. In fact, somewhat surprisingly, including the area does not substantially increase the survey c.v.s (Table 2).

Table 2: Coefficients of variation (c.v.s) of Atka mackerel biomass estimates from the trawl survey, by area, and overall (Aleutian Islands + Bering Sea). [The first two lines of this table are directly from table 15.4 of Lowe et al. (2007); the third line was calculated as $[(c_1B_1)^2 + (c_2B_2)^2]^{0.5}/(B_1 + B_2)$, where c_i and B_i are the c.v. and biomass estimate, respectively, from the ith area.]

	Coefficients of variation (%)						
Area	1991	1994	1997	2000	2002	2004	2006
Aleutian Islands	15	33	29	28	20	17	28
Bering Sea	37	99	99	87	99	43	44
Overall	15	31	31	28	20	16	28

3.2.3 Other issues

It didn't make sense to me that the *1986 survey biomass* estimate was used in the assessment (even with a high c.v.) when it was clearly stated that this was 'not directly comparable' with the subsequent surveys 'because of differences in the net, fishing power of the vessels, and sampling design' (p. 834, Lowe at al 2007). I have slightly less concern about using the at-age data from the 1986 survey, but even here, I would have thought that the differences mentioned in the above quote could substantially change the selectivity.

The issue of *spatial structure* was much discussed during the review, with the strongest information being the between-area differences in length frequencies (figure 15.8, Lowe et al 2007) and mean lengths at age (Lowe et al. 1998). The differences in length frequencies were so large that it seemed unlikely to me that they were caused solely by different growth rates. That is, there are likely to be between-area differences in age structure. The question is, how best to use this information in the assessment? I am doubtful as to the wisdom of constructing a full spatially-structured model, in which there is age- or length-dependent movement and differential recruitment between areas. It seems unlikely, at least in the short term, that there would be sufficient information to reliably estimate enough parameters to make this model clearly superior to the present one. My suggestion is to try to incorporate the spatial information by using area-specific selectivities for survey and fishery. That means having a separate fishery for each area. Also, use area-specific weight-at-age relationships in the model equation that converts catch weights to numbers of fish. This results in model which keeps track of biomass reasonably well, even though the numbers of fish may not quite make sense. This might sound odd, but note that all key model outputs are in terms of biomass rather than numbers.

Another topic of discussion was the potential to use *tagging-derived data in the assessment* (particularly abundance & movement estimates from McDermott et al. 2005). I think the currently available estimates are not useful in this way because, by design, they refer to spatial scales that are too small. It may be possible to use existing and future returns from this study to make preliminary estimates of rates of movement over larger distances. However, in the context of the stock assessment, I think the main value of this experiment is as a pilot study for a larger-scale experiment aimed at estimating abundance for the whole stock. Given the limited usefulness of the trawl survey in this assessment, a new fishery-independent method of estimating stock size would be of great value.

I suggest reducing the *plus group* in the at-age data to 11 y or 12 y on the grounds that there is no significant information about year-class strengths in the proportions at older ages in these data sets (see figs 15.18 & 15.19, Lowe et al. 2007).

3.3 Pollock assessment

3.3.1 Survey catchability

In this assessment, survey catchability, q^s, was fixed at 1.0, on the grounds that this would be conservative (because the true value of q^s would almost certainly be less than 1, and so the true current biomass would be higher than estimated). This decision was based on analyses carried out in the 2004 assessment, in which the effects of estimating q^s were investigated. What these analyses showed is that when q^s was estimated, the estimate was much less than 1.0 (0.06), and all data sets were much better fitted (cf Models 1 and 2 in table 18 of Barbeaux et al 2004). I expected the lower estimate of q^s and the improved fit to the survey biomass, because these are consistent with the simple 'no-response' hypothesis I described above (see Section 3.1.1). What I did not expect, and what seems inconsistent with this simple hypothesis, is the improved fit to the two at-age data sets (which was confirmed during the assessment meeting when a profile on q^s was constructed). I feel this needs more exploration. One possibility, mentioned during the review (but not, I think, in the assessment report), is that adult fish from other stocks migrate into the area, thus reducing apparent mortality and suggesting higher biomass. Another possibility is that the high values of fishing mortality estimated for 1995-1998 (see table 1A.23 in Barbeaux et al 2007) are inconsistent with the catch-at-age data, and that reducing q^s lessens this inconsistency. If this were true, then it may be possible to find a lower value of q^s that is still conservative, but more consistent with the at-age data. We would still have to fix q^s in the assessment, but the fixed value would derive from the data. This latter possibility could be investigated by disaggregating the at-age data likelihoods (i.e., calculate L_{tay}, the negative log-likelihood for the proportion at age *a* in year *y* from data set *t* – either survey or catch) and seeing which individual observations are better fit as q^s decreases (this analysis needs some imaginative use of graphics).

3.3.2 Survey selectivity

I think it was probably a mistake to allow the survey selectivity to vary from year to year. The reason given for allowing this variation was that a discussion with the people who have run this survey elicited comments on a number of changes of survey practice that could have changed the survey selectivity from year to year. I accept that this is a good reason to *investigate* changes in selectivity. However, we should allow these changes in the model only if there is clear evidence that they are estimable from the assessment data (note that although we have very good reason to believe that natural mortality, *M*, varies with time, we rarely estimate time-varying *M* because our data do not allow us to reliably distinguish this from other variation). A quick investigation, during the review, suggested to me that it was not worthwhile to allow time-varying survey selectivity. However, this decision was not clear-cut, for reasons that are discussed above in Section 3.1.3.

3.3.3 Assessment area

A feature of this assessment was that the eastern part of the assessment area (ENRA) was excluded for catches but included for survey data (this was so in models 2A, 2B, and 2C; model 1 was slightly different). I find this unusual, but I can conceive of situations in which it would be a reasonable thing to do. What seemed lacking to me in the assessment report was a rigorous conceptual model that justified the decision to treat the data in this way. The model, as presented during the review, concerned seasonal movements of fish from other stocks into and out of ENRA. This hypothesis seemed to support the treatment in the assessment of the survey data and recent catches, but I wasn't sure that it was consistent with the treatment of earlier catches (which occurred at a different time of year – see slide 12, 'Stock Assessment overview_pollock.ppt').

On a related topic, there was some discussion as to whether it was appropriate to assess a stock that is fished in the winter using data from a summer survey. The discussion raised two possible objections to doing this. First, the survey may not be sampling the same population, because of seasonal movements in and out of the survey area. This is clearly a subject of concern that can only be resolved by obtaining seasonal movement information to test the conceptual model discussed in the previous paragraph. The second possible objection is that the fish may be behaving very differently (perhaps in terms to their depth distribution) in the different seasons. I think this is not likely to be a problem, and will be easily dealt with by differences in survey and fishery selectivities (unless, of course, the difference in distribution is such that a large proportion of the population is not available to the survey).

3.3.4 Model selection

I thought the reasons given for model selection (p. 148, Barbeaux et al. 2007) weren't very good. First, it was stated that model 1 was rejected because the other models had lower 'quasi-likelihood'. This is an inappropriate comparison because likelihoods are not comparable when different data sets are used. What I would have like to have seen is a plot of the fit to the survey biomass estimates in model 1. I suspect that this would have allowed us to reject this model on the grounds that it did not mimic the substantial drop in biomass estimate in the last two years of the survey. The next comparison was between models 2A and 2B, which were identical, except that one more parameter, natural mortality, was estimated in the latter. It was said the Model 2B ‘improves the model fit over Model 2A’. This couldn’t be judged from the likelihood table in the assessment report (which was garbled), but it was certainly not true in the table presented during the review (see slide 32, ‘Stock Assessment overview_pollock.ppt’). Here, the fit to the data sets was actually slightly worse in 2B, by 0.26 points (102.90 – 102.64), and the overall fit was only slightly better (by 0.21 = 114.14 – 113.93) for 2B – not enough to justify the extra parameter.

Finally, Model 2C was said to be worse than 2B because it ‘degrades the overall fit to the catch-at-age data’. Well, it’s true that the fit to the data is worse, but only very slightly – by 1.09 points (103.99 – 102.90). What I think more important when comparing 2B and 2C is to consider the difference in their assumptions. This had something to do with ‘Modeling recruitment from 1978–2007 in Model 2C’, but it was difficult to find out exactly what the difference was. What I did find out is that the very large 1978 year class was excluded (in some sense) from Model 2B, apparently on the grounds that it might have been atypical. That doesn’t seem sensible to me. I can imagine having a base model with includes this year class, and another model that excludes it as a sensitivity analysis, but I heard no good reason to exclude it from the base model.

3.3.5 Other issues

It doesn’t seem sensible that the *mean of the prior on M* (0.3, in Models 2B and 2C) should differ from the value that was used when this parameter was fixed (0.2, in Models 1 and 2A).

I was surprised by the choice of *fishery selectivity for use in projections*. Normally, I would expect that the estimated selectivity from the most recent year (2007) would be used. The 2007 selectivity was rejected on the very reasonable grounds that it was atypical, because there hasn’t been a proper fishery since 1998. However, this selectivity appears very similar to that estimated for the early 1990s, before the fishery was effectively closed (see fig. 1A.23, Barbeaux et al. 2007), and is very different from the selectivity that was chosen for the projections (slide 40 in ‘Stock Assessment overview_pollock.ppt’).

4. Conclusions

The following conclusions are organised according to the two sets of bullet points provided in the Terms of Reference section of the Statement of Work (see Appendix 1).

4.1 Overall strengths and weaknesses

In addition to the four bullet points under this heading in the Terms of Reference I have added an additional one on documentation.

4.1.1 Analysts' use of data

The analyst's understanding and use of the available data were generally good. My only concerns related to the choice of areas for some data sets in both assessments (Sections 3.2.2, 3.3.3) and the use of 1986 survey data for Atka mackerel (first paragraph, Section 3.2.3).

4.1.2 Modelling methodology and logic

I found a number of areas in which I think the modelling methodology could have been a little better. However, I stress that the assessments seemed basically sound, so that the effect of changes in these areas would be incremental, rather than substantial.

The areas of concern for both assessments related to data weighting (Section 3.1.2), selectivity smoothers (Section 3.1.3) and AMAK model diagnostics (Section 3.1.5). For the pollock assessment, I had concerns relating to the survey selectivity (Section 3.3.2), the logic of model selection (Section 3.3.4), and two other issues (Section 3.3.5). In the Atka mackerel assessment there was the plus group (last paragraph, Section 3.2.3).

4.1.3 Uncertainties

In both assessments the statistical uncertainties were calculated using standard techniques and were well described. However, the considerable additional uncertainty associated with survey catchability did not seem to me to have been well communicated, although I am unsure about local conventions as to how and where such uncertainty should be expressed. I also felt that more could have been done to explore the causes and nature of this uncertainty (Section 3.1.1).

The question as to whether uncertainties were 'appropriately applied to management advice' is a thorny one because what is deemed appropriate seems to vary hugely amongst different fisheries administrations. In the current context I don't feel I have sufficient information to have an opinion on this question.

4.1.4 Best available science

What is the best available science for stock assessments is constantly evolving and, in some areas, a subject of some debate. I found no areas in which the current assessments departed substantially from best practice, but I have some suggestions for modifications to AMAK that might provide incremental improvements to future assessments (Section 3.1.5).

For Atka mackerel, there are two areas with potential for future improvement (Section 3.2.3). One, relating to spatial structure, could be explored with current resources; the other, the use of tagging data, would require additional resources.

4.1.5 Documentation

I think there is substantial scope for improvement in the documentation of both the assessments and the associated software (Section 3.1.6).

4.2 Additional specific topics

The determination of appropriate sample size for the multinomial distribution used for survey and fishery catch-at-age in both models. Suggestions are provided in Sections 3.1.2 and 3.1.5, and also Appendix 5.

The incorporation of differential growth parameters for Atka mackerel. I suggest an approach part-way between the current model and one that is fully spatially-structured (second paragraph, Section 3.2.3).

The incorporation of abundance and movement information from tagging studies of Atka mackerel. Tagging abundance estimates could greatly enhance the Atka mackerel assessment, but this would need a new purpose-designed tagging experiment (third paragraph, Section 3.2.3).

The potential pitfalls and possible solutions to the use of pollock summer bottom-trawl abundance index for a fishery that primarily occurs in the winter on a pelagic spawning population. This is clearly a subject of concern that can only be resolved by obtaining seasonal movement information to test the conceptual model currently used in the assessment (second paragraph, Section 3.3.3).

For pollock assess the appropriate spatial delineation of fisheries and survey data. The conceptual model underlying this delineation appears to have both spatial and temporal (seasonal) components, but needs to be more clearly specified and, if possible, tested (first paragraph, Section 3.3.3).

5. References

Akaike, A. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC-19*: 716-723.

Barbeaux, S.; Ianelli, J.; Brown, E. (2004). Stock Assessment of Aleutian Islands Region Pollock.

Barbeaux, S.; Ianelli, J.; Gaichas, S.; & Wilkins, M. (2007). Stock Assessment of Aleutian Islands Region Pollock.

Bull, B.; Francis, R.I.C.C.; Dunn, A.; McKenzie, A.; Gilbert, D.J.; Smith, M.H.; Bian, R. (2008). CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.20-2008/02/14. NIWA Technical Report 130. 275 p. [This document, and the associated program, can be obtained by anonymous ftp from ftp://ftp.niwa.co.nz/software/casal/latest_casal.zip].

Francis, R.I.C.C.; Hurst, R.J.; Renwick, J.A. (2003). Quantifying annual variation in catchability for commercial and research fishing. Fishery Bulletin 101: 293-304.

Kastelle, C.R.; Kimura, D.K. (2006). Age validation of walleye pollock (Theragra chalcogramma) from the Gulf of Alaska using the disequilibrium of Pb-210 and Ra-226. ICES Journal of Marine Science 63: 1520-1529.

Lowe, S.; Ianelli, J.; Wilkins, M.; Aydin, K.; Lauth, R.;, & Spies, I. (2007). Stock Assessment of Aleutian Islands Atka Mackerel.

Lowe, S.; Ianelli, J.; Zenger, H.; Aydin, K.; & Lauth, R. (2004). Stock Assessment of Aleutian Islands Atka Mackerel.

Lowe, S.A.; Van Doornik, M.D.; & Winans, G.A. (1998). Geographic variation in genetic and growth patterns of Atka mackerel, Pleurogrammus monopterygius (Hexagrammidae), in the Aleutian archipelago. Fisheru Bulletin 96: 502-515.

McDermott, S.F.; Fritz, L.W.; & Haist, V. (2005). Estimating movement and abundance of Atka mackerel (Pleurogrammus monopterygius) with tag-release-recapture data. Fisheries Oceanography: 14 (Suppl. 1): 113-130.

Plan Team (2007). Stock Assessment And Fishery Evaluation Report For The Groundfish Resources Of The Bering Sea/Aleutian Islands Regions. Compiled by

The Plan Team for the Groundfish Fisheries of the Bering Sea and Aleutian Islands.

APPENDIX 1: Statement of Work

Aleutian Islands Atka Mackerel and Pollock Stock Assessments

Panel Review Meeting 9-13 June 2008

General

The Alaska Fisheries Science Center (AFSC) requests a Center of Independent Experts (CIE) review of stock assessments for the Aleutian Islands stocks of Atka mackerel and pollock. In the Aleutian Islands Atka mackerel and pollock are key prey for several top trophic level consumers in the region. Of particular concern, Atka mackerel and pollock are dominant prey items for the endangered Steller sea lion. In addition, Aleutian Islands Atka mackerel supports a valuable commercial fishery. The pollock fishery was closed to directed fishing between 1999 and 2004 due to concerns for Steller sea lion recovery. Directed fishing is still restricted to outside of SSL critical habitat. A limited fishery outside SSL critical habitat was attempted in 2005, but resulted in very little catch (~200 t). In 2006 and 2007 a fishery within SSL critical habitat was conducted in conjunction with a cooperative acoustic survey under an exempted fishing permit, but total removals per year remained below 2,500 mt. There is a high level of interest from commercial fishers in reestablishing a directed pollock fishery in the Aleutian Islands. Because of their unique role in the Aleutian Island ecosystem and their importance to industry, it is critical that biomass is estimated accurately and that harvest recommendations are set in a manner that will sustain the resource and its predators. Both the pollock and Atka mackerel assessments utilize the same age-structured statistical model, and these species share many life history and population dynamics characteristics. Several changes have been made to improve the assessments and these changes have never been formally reviewed by a CIE panel. Several recent research projects have focused attention on the seasonal movements, stock structure and reproductive ecology of Atka mackerel and pollock. We will be seeking advice on techniques to incorporate this information into the assessment.

Overview of CIE Peer Review Process:

The Office of Science and Technology implements measures to strengthen the National Marine Fisheries Service's (NMFS) Science Quality Assurance Program (SQAP) to ensure the best available high quality science for fisheries management. For this reason, the NMFS Office of Science and Technology coordinates and manages a contract for obtaining external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of stock assessments and various scientific research projects. The primary objective of the CIE peer review is to provide an impartial review, evaluation, and recommendations in accordance to the Statement of Work (SoW), including the Terms of Reference (ToR) herein, to ensure the best available science is utilized for the National Marine Fisheries Service management decisions.

The NMFS Office of Science and Technology serves as the liaison with the NMFS Project Contact to establish the SoW which includes the expertise requirements, ToR, statement of tasks for the CIE reviewers, and description of deliverable milestones with dates. The CIE, comprised of a Coordination Team and Steering Committee, reviews the SoW to ensure it meets the CIE standards and selects the most qualified CIE reviewers according to the expertise requirements in the SoW. The CIE selection process also requires that CIE reviewers can conduct an impartial and unbiased peer review without the influence from government managers, the fishing industry, or any other interest group resulting in conflict of interest concerns. Each CIE reviewer is required by the CIE selection process to complete a Lack of Conflict of Interest Statement ensuring no advocacy or funding concerns exist that may adversely affect the perception of impartiality of the CIE peer review. The CIE reviewers conduct the peer review, often participating as a member in a panel review or as a desk review, in accordance with the ToR producing a CIE independent peer review report as a deliverable. The Office of Science and Technology serves as the COTR for the CIE contract with the responsibilities to review and approve the deliverables for compliance with the SoW and ToR. When the deliverables are approved by the COTR, the Office of Science and Technology has the responsibility for the distribution of the CIE reports to the Project Contact.

Requirements for CIE Reviewers:

The CIE assessment review requires a total of three CIE reviewers who are thoroughly familiar with various subject areas involved in stock assessment, including population dynamics, separable age-structured models, harvest strategies, survey methodology, and the AD Model Builder programming language. They should also have experience conducting stock assessments for fisheries management. Three CIE reviewers are requested to conduct an impartial and independent peer review in accordance with the Terms of Reference (ToR) herein. Each CIE reviewer's duties shall not exceed a maximum of 14 days conducting pre-review preparations with document review, participation in the panel review meeting, and completion of the CIE independent peer review report in accordance with the ToR and Schedule of Milestones and Deliverables.

Specific Activities and Responsibilities

Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the Terms of Reference (ToR) herein. The reviewers will travel to Seattle, Washington, to participate during a panel review meeting on AFSC's Atka mackerel and pollock stock assessment, conduct the independent peer review, and provide editorial assistance to the Chair with the summary report. Overview presentations by AFSC scientists will be made on several topics to facilitate the review, and assessment authors will be available for questions from reviewers.

Prior to the Peer Review: The CIE shall provide the CIE reviewers contact information (name, affiliation, address, email, and phone), including information needed for foreign travel clearance when required, to the Office of Science and Technology COTR no later than the date as specified in the SoW. The Project Contact is responsible for the completion and submission of the Foreign National Clearance forms (typically 30 days before the peer review), and must send the pre-review documents to the CIE reviewers as indicated in the SoW.

Foreign National Clearance: If the SoW specifies that the CIE reviewers shall participate in a panel review meeting requiring foreign travel, then the CIE shall provide the necessary information (e.g., name, birth date, passport, travel dates, country of origin) for each CIE reviewer to the COTR who will forward this information to the Project Contact. The Project Contact is responsible for the completion and submission of required Foreign National Clearance forms with sufficient lead-time (30 days) in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations at the Deemed Exports NAO link http://deemedexports.noaa.gov/sponsor.html

Pre-review Documents: Approximately two weeks before the peer review, the Project Contact will send the CIE reviewers the necessary documents for the peer review, including supplementary documents for background information. The CIE reviewers shall read the pre-review documents in preparation for the peer review.

Each of the reviewers shall generate individual reports. In addition, the chairperson shall generate a Summary Report that compiles the points made by the three individual reviewers into one succinct document. The individual reports shall be appended to the Summary Report, thereby providing the complete detailed information from the individual reviewers.

Terms of Reference

All reports shall address the following points.

- The strengths and weaknesses of the modeling efforts for Aleutian Islands Atka mackerel and pollock assessments and harvest recommendations. Specifically, the review shall evaluate:
    - The analysts' use of fishery dependent and fishery independent data sources in the assessments;
    - Gaps or inconsistencies in the population dynamics modeling methodology or logic;
    - If uncertainties in assessment model results are appropriately applied to management advice; and
    - Whether the assessments provide the best available science.

Additionally, the review shall (to the extent practical) evaluate and provide advice on:

- The determination of appropriate sample size for the multinomial distribution used for survey and fishery catch-at-age in both models.
- The incorporation of differential growth parameters for Atka mackerel
- The incorporation of abundance and movement information from tagging studies of Atka mackerel
- The potential pitfalls and possible solutions to the use of pollock summer bottom-trawl abundance index for a fishery that primarily occurs in the winter on a pelagic spawning population.
- For pollock assess the appropriate spatial delineation of fisheries and survey data.

The AFSC will provide copies of stock assessment documents, survey reports, and other pertinent literature on a web site.

Specific

1. Read and become familiar with the relevant documents provided to the reviewers.

2. Discuss the stock assessment with the lead assessment scientist and survey scientists in Seattle, Washington, from June 9 to June 13, 2008.

3. No later than June 27, 2008, submit a written report of findings, analysis, and conclusions. More details on the report outline and organization are provided in Annex I.

Schedule of Milestones and Deliverables

The milestones and schedule are summarized in the table below. No later than June 27, 2008, the CIE panelists should submit their CIE independent peer review reports to the CIE for review[1]. These reports shall be submitted to Mr. Manoj Shivlani, CIE Lead Coordinator, via email at shivlanim@bellsouth.net, and to Dr. David Die, CIE Regional Coordinator, via email at ddie@rsmas.miami.edu.

Milestone	**Date**
CIE will provide CIE reviewer contact information, and project contact will distribute pre-meeting material to the CIE reviewers	May 26, 2008
CIE reviewers attend the **Atka Mackerel and Pollock Stock Assessment meeting** to conduct peer review at AFSC, Seattle, WA, USA	June 9-13
CIE reviewers submit CIE independent peer review reports to CIE for approval	June 27
CIE provides reviewed CIE independent peer review reports to NMFS COTR for SOW and ToR compliance approval	July 3
COTR notifies CIE of approval of CIE independent peer review reports	July 4
COTR provides final CIE independent peer review reports to AFSC contact	July 5

Acceptance of Deliverables:

Upon review and acceptance of the CIE reports by the CIE Coordination and Steering Committees, CIE shall send via e-mail the CIE reports to the COTRs (William Michaels William.Michaels@noaa.gov and Stephen K. Brown Stephen.K.Brown@noaa.gov) at the NMFS Office of Science and Technology by the date in the Schedule of Milestones and Deliverables. The COTRs will review the CIE reports to ensure compliance with the SoW and ToR herein, and have the responsibility of approval and acceptance of the deliverables. Upon notification of acceptance, CIE shall send via e-mail the final CIE report in *.PDF format to the

[1] All reports will undergo an internal CIE review before they are considered final.

COTRs. The COTRs at the Office of Science and Technology have the responsibility for the distribution of the final CIE reports to the Project Contacts.

Key Personnel:

Contracting Officer's Technical Representative (COTR):
William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Stephen K. Brown
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Stephen.K.Brown@noaa.gov Phone: 301-713-2363 ext 133

Contractor Contacts:
Manoj Shivlani, CIE Lead Coordinator
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Project Contact:
Sandra Lowe
Sandra.Lowe@noaa.gov, Phone: (206) 526-4230;

Steve Barbeaux,
Steve.Barbeaux@noaa.gov, Phone: (206) 526-4211

Request for Changes:

Requests for changes shall be submitted to the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the Contractor within 10 working days after receipt of all required information of the decision on substitutions. The contract will be modified to reflect any approved changes. The Terms of Reference (ToR) and list of pre-review documents herein may be updated without contract modification as long as the role and ability of the CIE reviewers to complete the SoW deliverable in accordance with the ToR are not adversely impacted.

Annex 1: Contents of the Reviewer's Report

The following requirements refer to all reports, both the individual reports and the Summary Report.

1. All reports shall be prefaced with an executive summary of findings and/or recommendations.
2. In accordance with each Term of Reference, the main body of all reports shall consist of a background, description of review activities, summary of findings, conclusions and recommendations, and references.
3. The reports shall also include as separate appendices, the bibliography of all materials provided and any additional papers cited, along with a copy of the statement of work and meeting agenda.

APPENDIX 2: Materials Provided

Before the review the Panel was provided with two groups of documents via the site ftp://ftp.afsc.noaa.gov/afsc/public/atka/default.htm. The first and most important group is described below. For the second group (Background documents), and a third group provided during the course of the review (Presentations to the CIE Review Panel) see the above site.

Assessment Reports

Lowe, S.; Ianelli, J.; Wilkins, M.; Aydin, K.; Lauth, R.;, & Spies, I. (2007) Stock Assessment of Aleutian Islands Atka Mackerel

Barbeaux, S.; Ianelli, J.; Gaichas, S.; & Wilkins, M. (2007) Stock Assessment of Aleutian Islands Region Pollock

A link to all Alaska groundfish stock assessments http://www.afsc.noaa.gov/refm/stocks/assessments.htm

Reports from earlier assessments of Atka mackerel (2006, 2004, 2002) and pollock (2006, 2005, 2004, 2003)

Plan Team (2007). Stock Assessment And Fishery Evaluation Report For The Groundfish Resources Of The Bering Sea/Aleutian Islands Regions. Compiled by The Plan Team for the Groundfish Fisheries of the Bering Sea and Aleutian Islands.

2008 - 2009 NPFMC TAC/ABC Recommendations

APPENDIX 3: List of Participants

Participants in the review meeting included the following

Review Panel

Chris Francis, Ana Parma, Kurtis Trzcinski

Assessment team

Steve Barbeaux, Jim Ianelli, Sandra Lowe

Others

Ann Hollowed (introduction); Lisa Thompson & Jennifer Cahalan (observer sampling); Delsa Andrel & Betty Goetz (age & growth); Dan Kimura (age validation); Mark Wilkins (trawl survey); Ingrid Spies (mackerel genetics); Dan Cooper (mackerel ecology); Suzanne McDermott (mackerel tagging); Kerim Aydin (ecosystem modelling); Doug Kinzey (multispecies modelling)

APPENDIX 4: Agenda

Monday June 9th

9:00	Welcome and Introductions	
9:15	**Overview** (management, fishery, biology descriptions)	
	Management control rules and general modeling approach	**Jim**
	Atka mackerel	**Sandra**
	Pollock	**Steve**
11:30	**Observer** sampling and coverage	**Lisa Thompson and Jennifer Cahalan**
12:00	**Lunch**	
13:00	Age and growth	**Delsa Anderl and Betty Goetz**
13:30	Bottom trawl survey	**Mark Wilkins**
	Research	
14:15	Atka mackerel tagging	**Susanne McDermott**
14:30	Spawning characteristics and habitat for Atka mackerel	**Bob Lauth**
14:45	Genetics	**Ingrid Spies and Mike Canino**
15:00	Cooperative research survey on pollock	**Steve**
15:30	Aleutian Islands **ecosystem** overview	
	FEP, foodweb linkages	**Kerim Aydin**
16:00	Age-structured multispecies modeling	**Doug Kinzey**

Tuesday June 10th

Atka mackerel and pollock stock assessments

9:00	Assessment model details	**Jim**
10:00	Atka Mackerel stock assessment	**Sandra**
10:45	Break	
11:00	Atka Mackerel stock assessment (continued)	**Sandra**
12:00	**Lunch**	
13:00	Pollock stock assessment	**Steve**
14:45	Break	
15:00	Stock assessment issues	
	Initial age composition, recruitment, effective N, incorporation of uncertainty, selectivity, stock-recruitment relationships	
	Reviewer discussions with assessment authors	

Wednesday June 11th

Reviewer discussions with assessment authors

APPENDIX 5: Multinomial sample sizes

This Appendix describes two methods of correcting the multinomial sample sizes for at-age observations in a stock assessment model. With both methods we run the assessment model with initial sample sizes, and then analyse the residuals from the model fit to calculate a correction factor for the sample size. The aim is to ensure that the size of the model residuals is consistent with the corrected sample sizes.

Let $p_{ay,\mathrm{obs}}$ be the observed proportion at age a in year y in a set of age frequencies that is assumed to have a multinomial error structure, and let $N_{\mathrm{init},y}$ denote the initial sample sizes. Our aim is to calculate a correction factor f so that the size of the model residuals is consistent with the corrected sample sizes, $N_{\mathrm{corr},y} = fN_{\mathrm{init},y}$.

There two ways of calculating f, and these use different residuals. The first, and most usual method, uses the residuals of the individual observations: $r_{ay} = p_{ay,\mathrm{obs}} - p_{ay,\mathrm{exp}}$; the second method uses the residuals of mean age: $r_y = (m_{y,\mathrm{obs}} - m_{y,\mathrm{exp}})$, where $m_{y,\mathrm{obs}} = \Sigma_a(a\, p_{ay,\mathrm{obs}})$, and $m_{y,\mathrm{exp}}$ is defined similarly. I now describe the two methods.

For the first method, we use the fact that the expected variance of $p_{ay,\mathrm{obs}}$, and thus of r_{ay}, should be $t_{ay}/(fN_{\mathrm{init},y})$ [i.e., $t_{ay}/N_{\mathrm{corr},y}$], where $t_{ay} = p_{ay,\mathrm{exp}}(1 - p_{ay,\mathrm{exp}})$. Therefore, the expected variance of $r_{ay}(N_{\mathrm{init},y}/t_{ay})^{0.5}$ is $1/f$, and we can estimate f as $1/\mathrm{Var}(r_{ay}(N_{\mathrm{init},y}/t_{ay})^{0.5})$.

For the second method, the expected variance of the mean length, $m_{y,\mathrm{obs}}$, and thus of the residual r_y, is $v_y/(fN_{\mathrm{init},y})$ [i.e., $v_y/N_{\mathrm{corr},y}$], where v_y is the variance of the age frequency in year y, given by $v_y = \Sigma_a(a^2\, p_{ay,\mathrm{obs}}) - m_{y,\mathrm{obs}}^2$. Therefore, the expected variance of $r_y(N_{\mathrm{init},y}/v_y)^{0.5}$ is $1/f$, and we can estimate f as $1/\mathrm{Var}(r_y(N_{\mathrm{init},y}/v_y)^{0.5})$.

I recommend using the second of these methods, which will often produce a smaller (sometimes much smaller) value of f. It is smaller because there are often strong correlations between the age proportions in the same year (i.e., between p_{ay} and $p_{a'y}$) which arise because two fish from the same catch are typically more alike (in size and/or age) than two fish from different catches. With these correlations the first method is invalid. However, it should be noted that the second method will not be very accurate when the at-age data cover few years (much fewer than 10 years, say).

Postscript

The above methods assume that the best way to correct the sample sizes is with a multiplicative factor f so that $N_{\text{corr},y} = fN_{\text{init},y}$. A case can be made that a better equation is $1/N_{\text{corr},y} = (1/N_{\text{init},y}) + (1/N_{\text{adj}})$. Here, our aim would be to estimate N_{adj} rather than f. It wouldn't be difficult to construct an estimator for N_{adj}. I've not done that because (a) the multiplicative factor seems to be the more common approach in Alaskan assessments, (b) f is more easily interpretable than N_{adj}, and (c) I'm concerned about robustness in the estimation of N_{adj}.