Report on the 2009 assessments of Bocaccio and Widow Rockfish off the U.S. West Coast

NIWA Client Report: WLG2009-54

NIWA Project: NTV10301

Report on the 2009 assessments of Bocaccio and Widow Rockfish off the U.S. West Coast

R.I.C.C. Francis

NIWA contact/Corresponding author

Prepared for

The Center for Independent Experts

Northern Taiga Ventures, Inc.

NIWA Client Report: WLG2009-54

NIWA Project: NTV10301

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

© All rights reserved. This publication may not be reproduced or copied in any form without the permission of the client. Such permission is to be given only in accordance with the terms of the client's contract with NIWA. This copyright extends to all forms of copying and any storage of material in any kind of information retrieval system.

Contents

Executive Summary		iv
1.	Background	1
2.	Review Activities	1
3.	Findings	1
3.1	Findings common to both species	1
3.2	Findings for bocaccio	6
3.3	Findings for widow rockfish	9
4.	Major conclusions and recommendations	10
5.	References	11
APPENDIX 1: Materials Provided for the Review		12
APPENDIX 2: Statement of Work		13
APPENDIX 3: Panel Membership		21
APPENDIX 4: Multinomial sample sizes		22

Approved for release by:

Dr Andrew Laing

Executive Summary

A panel reviewed the 2009 assessments of bocaccio and widow rockfish stocks off the U.S. west coast. The Panel met 13-17 July 2009 at the Fisheries Ecology Division of NOAA/NMFS in Santa Cruz. The draft assessments, together with related material, were presented to the Panel, additional analyses were requested and carried out, and the Panel discussed the results and drafted their report.

Both assessments were shown to be sensitive to the weightings assigned to different data sets. The approach to data weighting that was used in both draft assessments (and which seems to be fairly standard in recent west-coast U.S. assessments) tends to give too much weight to composition data, fails to ensure that at least some biomass indices are adequately fitted, and ignores expert opinion about the relative plausibility of each biomass index. Ways in which this approach might be improved are discussed.

The bocaccio assessment, as modified during the review, represents best available science. Further consideration of the weighting the composition data is recommended.

The final widow rockfish assessment was a clear improvement on the draft assessment, though there was insufficient time for the Panel to fully evaluate it. Also, it was unclear whether a simpler one-area model, which may well be superior, would produce a significantly different assessment. For these reasons it is not clear that the final assessment represents best available science. Evaluation of a one-area model is recommended, as is a reconsideration of the reliability of age data and of the triennial survey length data.

Further evaluation of Alec McCall's proposed delta method, as well as the complex recruitment bias adjustments included in Stock Synthesis 3, is recommended. For those biomass indices constructed using various types of GLM, investigations of alternative model assumptions (as was done for the CalCOFI survey) are worthwhile, but time-consuming Bayesian estimation is not.

1. Background

This report reviews, at the request of the Northern Taiga Ventures, Inc. (see Appendix 2), the 2009 assessments of bocaccio and widow rockfish stocks off the U.S. west coast. The author was provided with various documents (Appendix 1), and participated in the meeting which considered the assessment and in the writing of the Panel Report from that meeting.

2. Review Activities

The stock assessment review (STAR) panel met 13-17 July 2009 at the Fisheries Ecology Division of NOAA/NMFS in Santa Cruz. Those attending the meeting included four reviewers, three advisors, and the stock assessment teams (STATs) (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 STATs and drafted their report.

3. Findings

My findings in this section are in three groups: one for each species assessed, preceded by a group that was common to both species. Within these groups, the findings are organised according to the Terms of Reference (TOR) in Annex 2 of Appendix 2.

3.1 Findings common to both species

3.1.1 TOR 2: Data collection and survey design

Several survey data sets were common to both assessments (the triennial and NWFSC bottom trawl surveys, and the midwater trawl survey for juveniles). The time available for the review did not allow for a thorough examination of either data collection methods or of survey design. However, it was noted that such surveys, being designed to sample a broad suite of species, are not well suited to some individual species. The exclusion of the NWFSC survey from the Cowcod Conservation Areas (CCAs) is acknowledged to compromise its ability to index bocaccio (which are abundant in the CCAs); I was unclear as to whether this was a serious issue for widow rockfish.

3.1.2 TOR3: Data quality

In recent years fisheries for both species have undergone major changes because of management measures designed to protect the species and promote rebuilding.

Unfortunately, these changes may compromise fishery-dependent indices based on catch per unit effort (CPUE). The problem is that major management changes tend to induce major changes in the pattern of fishing, and these changes may affect catchability (the proportion of the population caught by one unit of effort) in ways that are not easily quantified. Both assessments assumed that catchability has not changed.

3.1.3 TOR4: Analytic methodologies

For several survey series (the triennial and NWFSC trawl surveys, the NWFSC hook and line survey, and the CalCOFI larval survey) the approach to the calculation of biomass indices used in these assessments was model-based (derived from generalised linear models (GLMs) of various types), rather than design-based (i.e., based on swept-area expansions). I am sympathetic to the use of this approach because it substantially reduces the effect of occasional large catches, and thus produces more credible indices. A disadvantage is that the c.v.s (coefficients of variation) produced by these models are unrealistically low. I have no solution to that problem. However, I am dubious about wisdom of fitting these models using very time-consuming Bayesian methods (both STAT teams referred to posterior sample sizes of more than one million). It is not that I have any reason to believe these methods to be wrong. It's simply that, given the quality and quantity of the data being analysed (i.e., bearing in mind that these surveys are not believed to index bocaccio or widow rockfish well), and given also that the output c.v.s are implausible, the application of Bayesian methods to produce these biomass indices and c.v.s seems like a very elaborate and time-costly way of putting lipstick on a pig.

I was impressed by the effort and thoroughness that went into the reconstruction of early catches from California (Ralston et al. 2009). Both the problem, and the available data sets, is complex, and I lack sufficient knowledge of these to be certain about the results. However, the brief description the Panel was given of the analytical approach gave me some confidence that the reconstructed catch histories are a great improvement on previous data.

3.1.4 TOR5: Model assumptions and sources of uncertainty

My main concern under this TOR was the procedure used by both STATs to 'tune' their models by iteratively reweighting the data sets. I understand this procedure to be fairly standard (perhaps with some minor variants) amongst west-coast U.S. assessment scientists. It involves assigning initial weights to each data set (in the form of c.v.s for biomass indices and multinomial sample sizes for composition data); fitting the model and calculating, from model residuals, adjustments to these weights

(in the form of offsets to add to the c.v.s, and multipliers for the sample sizes); then adjusting the weights and re-running the model. The procedure may then be iterated until the adjusted weights don't change much between iterations.

The question of data weighting is clearly very important. For both assessments reviewed here, our view of the current status of the stock (as expressed by the estimated depletion) was strongly dependent on the weightings given to different data sets. Moreover, these weightings strongly affect our interpretation of many standard diagnostics. They directly affect estimated confidence intervals, the shape and significance of likelihood profiles, and the use of AIC (Akaike 1974) to determine whether the gain in goodness of fit in an alternative model is justified, given the number of additional parameters that were used (e.g., choosing between the base case and alternative models in table 27 of the draft bocaccio assessment or table 14 of the draft widow assessment).

I have three concerns about the standard weighting procedure. The first is that it ignores valuable, though subjective, expert knowledge about the relative information content in different biomass indices. It effectively says to the model "here are all the data sets; do the best you can to fit them all, but give no precedence to any one data set, and don't worry if you can't fit some of them". If one biomass index is thought to be superior to others, then at least one model run should be done in which that index is sufficiently weighted that it is adequately well fitted (as a way of defining "adequately well fitted" I suggest that the model r.m.s.e. (root-mean square error) for the index should be similar to that achieved when a smooth line is fitted to it outside the model, and that there be no strong trend in the model residuals). When no index stands out as superior to the others I suggest an approach like that developed, during this review, for bocaccio. This involved grouping the indices into those that favour a more optimistic assessment, and those that are more pessimistic, and then making two bracketing runs: one that upweights the optimistic indices, and one that upweights those that are pessimistic.

My second concern relates to the relative weighting given to the two main types of data: biomass indices, and age and/or length compositions. The former type seems fundamentally more important to me because it relates directly to our prime concern in stock assessment: biomass trends (these trends *are* related to composition data, but only indirectly). Therefore I think our weighting system should give greater precedence to the biomass indices. We should expect our base case to fit at least some of the biomass indices "adequately well" (as defined above).

My third concern is that the standard weighting system gives too much weight to composition data because it ignores substantial correlations within those data. An alternative approach which allows for these correlations (see Appendix 4), was described during the review, applied to both assessments, and used in the final widow rockfish base case. This reduced the composition data sample sizes produced by the standard procedure by factors ranging from 0.1 to 0.4. Figure 1 provides an example of the scale of the correlations for one bocaccio data set (note that the multinomial error assumption used in the standard reweighting procedure implies these correlations should be very small – much less than 0.1 for this example – and negative). [This plot was not seen by the Panel – I didn't think of this method of demonstrating the existence of correlations until after the review.]

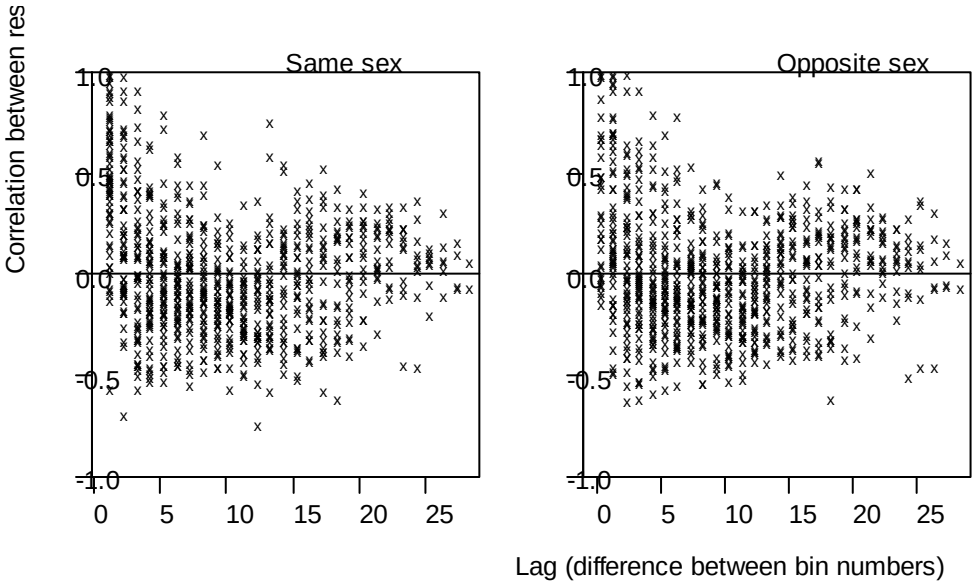

Figure 1: Correlations between residuals for bocaccio length composition data set trawlsou. Each plotted point is the correlation across years between residuals from two length-sex bins; the x-axis shows the 'lag' between the bins (i.e., the difference between the two bin numbers). For example, the top left point in the left-hand panel (a correlation of 0.9998) is the correlation between the proportions of females in the 3rd and 4th length bins.

3.1.5 TOR7: Further improvements

During the review, Alec McCall briefly described a method to include uncertainty about natural mortality and/or steepness in models in which these parameters are fixed. I support the Panel's recommendation that the applicability of this method should be evaluated. I also support another Panel recommendation concerning the recruitment bias adjustment scheme in Stock Synthesis 3 (SS3) that was used in both assessments. This is a complex and important matter than can have a significant effect

on an assessment. Thus it is important that the approach be thoroughly investigated and peer reviewed.

I have two small comments related to SS3 output used in iterative reweighting (see Section 3.1.4). This output is in the form of adjustments to the data weightings in the input files. It is potentially confusing to users (as was demonstrated during the review) that for biomass indices the output adjustments are intended to replace the input adjustments, whereas for composition data, they should be used together with any input adjustments (i.e., the new sample size multipliers should be the product of the input and output multipliers). Second, it seems wrong to me that the method of adjusting biomass index c.v.s should involve simple addition (i.e., if c_{iy} is the input c.v. for year y of the ith biomass index, then the adjusted c.v. is calculated as c_{iy} + $c_{i,\mathrm{adj}}$). The usual approach when combining two independent sources of error is add c.v.s as squares (i.e., $c_{i,\mathrm{adj}}$ should be calculated in such a way that the adjusted c.v. is $\left[c_{iy}^2 + c_{i,\mathrm{adj}}^2\right]^{0.5}$). This change is not simply computational. It would change the relative weighting between years within a biomass index.

I also have two small suggestions about the software that was used in both assessments to plot SS3 output. The first is that confidence intervals should be omitted in plots of fits to biomass indices if, as was often the case in these assessments, they expand the vertical scale so much that they obscure the trend in the data (e.g., the SCJuvSurvey index in the widow rockfish assessment drops by about 60% between the first and last four years, but this is hard to see in figure 25 in the draft assessment report). The second suggestion concerns a disjunction between the two types of plots involving spawning potential ratio (SPR) (e.g., figures E5, E6 in the draft bocaccio assessment report). It must be confusing to non-technical readers that high fishing pressure is represented in the first plot by points close to the x-axis, and in the second plot by points distant from the x-axis. This disjunction is easily removed if the y-axis in the first plot is modified to run from 1.0 to 0.0, rather than from 0.0 to 1.0.

3.1.6 TOR8: Review proceedings

My job as reviewer was greatly facilitated by the fact that both assessments used the same software (SS3), and that I was provided both with the SS3 documentation and the input files for the base case. My experience is that it can often be very difficult, and sometimes not possible, to answer, solely from the text of an assessment report, some fundamental questions about model assumptions (e.g., what parameters were estimated, and what error assumptions were used for each data set). For these assessments, it was easy to answer these questions.

Throughout the review meeting I was pleased to find the STATs keen to respond to questions, suggestions, and requests from the Panel. Fisheries Ecology Division were excellent hosts to the Panel, my only complaint being that the quantity and quality of food provided was more than was good for my waistline.

3.2 Findings for bocaccio

This assessment seemed very thorough to me. The STAT had carefully responded to suggestions from previous STAR panels, made a detailed examination of the transition from the previous assessment, and provided a wide array of diagnostics and sensitivity analyses. They responded quickly and thoroughly to all requests from the Panel.

3.2.1 TOR3: Data quality

I was pleased by the STAT's decision not to use age data in this assessment because they were unreliable. I do not know exactly how unreliable these data are, but I note that the length data were quite informative about strong and weak year classes, and I reject the view, sometimes expressed, that any data is better than no data. When otoliths are difficult to interpret there is a real danger that the interpretation will drift over time (or differ substantially between readers) thereby introducing misleading trends in age composition data. In New Zealand, we found this drift to be a particular problem with orange roughy (see section 2 of Francis 2006).

I noted two problems with the length composition data that reduce its reliability, and thus make it particularly important that these data are not over-weighted in the assessment (see Section 3.1.4). The data from the recreational fishery for some early years (1980–1992?) were recorded as weight frequencies and converted (outside the model) to length frequencies. Also, in the early years of the triennial surveys (1977–1986?) some smaller catches were not measured, and this introduced bias because fish in smaller catches tended to be larger than average.

I note also the improvement, since the last assessment, in the quality of the early part of the catch history, and the substantial effect that this had on estimated stock status (estimated depletion in 2006 changed from 13% to 27%).

3.2.2 TOR4: Analytic methodologies

I support the STAT's exploration of different link functions in the delta-GLM model used to derive biomass indices from the CalCOFI survey. This produced a markedly better fit to the binomial model, though the effect on the assessment was slight.

3.2.3 TOR5: Model assumptions and sources of uncertainty

Stock structure was a major source of uncertainty in this assessment. There were two key questions: where to place the northern and southern boundaries of the stock, and what to do about the apparent northward diffusion of fish as they age. Given the available data, the STAT's responses to these questions seemed sensible. Two sensitivity analyses were used to evaluate the effects of shifting the northern boundary, set at Cape Blanco in the base case, to the north or south. The diffusion problem was well dealt with via the use of different selectivities. A plot showing the average (across years) of each length composition data set (sexes combined) supported the use of domed selectivities for some data sets and un-domed selectivities for others.

Another important source of uncertainty was the appropriate weight to give to each of two pairs of biomass indices. One pair (the triennial survey and CPUE for the southern trawl fishery) showed a steep decline in biomass in the 1980s and supported a more pessimistic assessment. The other (the CalCOFI survey and CPUE from the southern recreational fishery) showed a strong rebuild in the early 2000s and supported a more optimistic assessment. The two bracketing runs in the final assessment are a good way of demonstrating this uncertainty. The base case provided a pragmatic, if somewhat arbitrary, middle way between these bracketing runs, but I found no reason to believe that it was any more plausible than either of the bracketing runs.

The selectivity curve for the triennial survey was found to be unstable, and so most of its parameters were fixed at values determined through jittering of early runs. It is not ideal to have to fix selectivity parameters in this way, but I acknowledge that it can be very difficult to resolve instability problems of this sort, and that fixing these parameters was a reasonable pragmatic response to the problem. It's possible this instability would disappear if the composition data were down-weighted (see Section 3.2.5).

3.2.4 TOR6: Best available science

I believe this assessment, as modified following suggestions from the Panel, represents the best available science.

3.2.5 TOR7: Further improvements

I am dubious about two possible improvements mentioned in the draft assessment report. The first, suggested by a previous STAR panel, was to seek correlations between environmental variables and the model residuals from the fit to the CalCOFI

index. This is, in principle, a good idea. A positive (or negative) residual in a particular year suggests that more (or fewer) larvae survived in that year, and it is of great interest to know what environmental conditions favour, or inhibit, larval survival. However, since the residuals are strongly dependent on the weightings applied to different data sets, and there is no compellingly obvious way to derive these weightings, it seems likely that much of the residual variation will be noise, rather than an indicator of year-to-year changes in survival.

The second possible improvement (mentioned by the STAT on pp. 10 & 62 of the draft assessment report) is to add area structure to the assessment model. It doesn't seem likely to me that there are sufficient data to inform such a model. This is not to say that there is not important spatial structure in this population. There clearly is (see Section 3.2.3). But there is little point in developing a more complex model if the available data will not allow good estimates of the additional parameters.

It was decided, during the review, not to use the method described in Appendix 4 to reweight the composition data on the grounds that its effect on the assessment was relatively small (it changed the estimated depletion from 22% to 19%) and the method was new. I suggest reconsidering this decision for the next assessment, particularly given the strong correlations shown in Figure 1 above (not seen by the Panel). As well as producing more reliable model diagnostics (see Section 3.1.4) this reweighting might reduce, or even remove, the above-mentioned instability of the selectivity for the triennial survey (when this reweighting was tried during the review it reduced the median sample size for the triennial length compositions to 2.9).

I was surprised that, in a model in which growth was estimated and the only composition data was for lengths, no parameters describing the variability in length at age (particularly for older ages) were estimated in the base case. These parameters can have a marked effect on the fit to the right-hand side of length frequencies. I understand that such parameters were estimated in earlier exploratory models, and presume that they were then simply fixed at estimated values for later models. I think that, unless they are causing stability problems, they are better kept as estimated. However, this is a pretty minor point.

I have two small suggestions that would be very helpful to reviewers like me, who do not have a deep familiarity with the local geography and data sets. All relevant geographical features should be included in a map (e.g., fig. 2 in the draft report (mislabelled as fig. 1) should have included Capes Blanco & Mendocino, Point Conception, and the boundaries for the two trawl fisheries and the separate north and south models). Also, be consistent with the labelling of data sets. What was called

NWFSCtrawl in SS3 was variously labelled 'Northwest Center Trawl Survey' (p.41), 'NWFSC combined trawl survey' (p. 53), 'NWFSC trawl survey' (p. 94), and 'Combo' (pp. 99-101). It also took me a while to make a connection between the labels CFGCPUE and CPFV, which referred to the same data.

3.3 Findings for widow rockfish

I found this a frustrating assessment to review. It was disappointing that there were so many problems with the catch and discard data, and with the data weightings (see next two sections). These problems took so much time to resolve that there was inadequate time to explore the new base case (which was quite different in its conclusions from the initial base case), let alone to consider sensitivity analyses.

3.3.1 TOR3: Data quality

During the review, several problems were identified with both the catch data (pre-1983 catches from the two Oregon fisheries were included in the Washington fishery, and the 1916–1976 catches for the latter fishery were incorrect) and the treatment of discards. I understand that these problems were resolved during the review.

There seems to be some doubt about the reliability of the age data. I was surprised to find that some of the early ages were from surface readings, as opposed to broken and burnt otoliths. Surface readings are generally considered unreliable for a relatively long-lived species such as this, except for the youngest ages. Also, the discussion during the review suggested that the precision of the break-and-burn ageing method is not well established for this species.

3.3.2 TOR5: Model assumptions and sources of uncertainty

During the review, problems were found with the data weightings in the draft base case model, so a new weighting scheme was applied. This substantially changed the perceived status of the stock.

I think that a major unresolved question for this assessment was whether the two-area model that was used was a significant improvement on a model without area structure (except in the form of area-specific fishery selectivities). The use of a two-area model required the already rather sparse data (both biomass indices and composition data) from each of two surveys (triennial and NWFSC) to be split into two parts, with no evaluation of whether this split was justified. The generic issue here is model complexity. The appropriate question to ask, when considering an increase in model complexity, is not whether the more complex model better matches reality. Our job in

stock assessment is *not* to mimic the real world; it is to make inferences about very specific aspects of the real world – primarily to infer the historical biomass trajectory of a fish stock. Thus the question to ask is, will a more complex model improve these inferences? In the present context, we would give a positive answer to this question only if (a) the two-area model produced a significantly different combined biomass trajectory than that from an initial one-area model, and (b) it was clear that this difference was driven by a real signal in the disaggregated data, and not just noise. In the two-area model there were six composition data sets (two for age, and four for length) associated with the two surveys that were split by area. For five of these data sets, the median adjusted sample sizes in the final base case model were less than 5. That suggests to me that there were insufficient data to justify the area split.

Another puzzling and unresolved issue is the fact that males had much greater mean length than females in catches from the north triennial survey despite the facts that (a) males were assumed to be smaller than females of the same age and (b) males were, on average, younger than females in catches from all three northern fisheries.

3.3.3 TOR6: Best available science

The agreed base case model at the end of the review was clearly a great improvement on that in the draft assessment report. However, I am not sure that it represents the best available science. Because of other problems described above, there was insufficient time to fully evaluate the final model. That might not have bothered me if, during the transition from the initial to the final base case, the estimated stock status (i.e., current depletion) had not changed as much as it did. Also, I am not sure that a one-area model, which could well be superior to the current base case, would not estimate a quite different stock status.

3.3.4 TOR7: Further improvements

The main improvement that I think is needed in this assessment is an evaluation of a one-area model, as discussed above. Before that is done, the reliability of the age data should be reconsidered and the triennial survey length and sex data examined (to seek an explanation for the apparent anomaly noted in Section 3.3.2).

4. Major conclusions and recommendations

I recommend a reconsideration of the approach to data weighting that was used in both draft assessments (I understand this approach to be fairly standard in recent west-coast U.S. assessments). This approach tends to give too much weight to composition data (this can be corrected using the method of Appendix 4), fails to ensure that at least

some biomass indices are adequately fitted, and ignores expert opinion about the relative plausibility of each biomass index.

I believe the bocaccio assessment, as modified during the review, represents best available science. My only major suggestion for its further improvement is that the methods of Appendix 4 be considered for weighting the composition data.

The final widow rockfish assessment was a clear improvement on the draft assessment, though there was insufficient time for the Panel to fully evaluate it. Also, it is unclear whether a simpler one-area model, which may well be superior, would produce a significantly different assessment. For these reasons I can not be confident that the final assessment represents best available science. As well as a one-area model, I recommend a reconsideration of the reliability of age data and of the triennial survey length data.

I also recommend further evaluation of Alec McCall's proposed delta method and the complex recruitment bias adjustments included in Stock Synthesis 3. For those biomass indices constructed using various types of GLM I suggest that investigations of alternative model assumptions (as was done for the CalCOFI survey) are worthwhile, but time-consuming Bayesian estimation is not.

5. References

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

Francis, R.I.C.C. (2006). Some recent problems in New Zealand orange roughy assessments. *New Zealand Fisheries Assessment Report 2006/43*. 65 p.

Ralston, S.; Pearson, D.E.; Field, J.C.; Key, M. (2009). Documentation of the California Catch Reconstruction Project

APPENDIX 1: Materials Provided for the Review

The Panel was provided with the following documents.

1. Draft stock assessment reports for bocaccio and widow rockfish.

2. Previous stock assessment reports for these species (2003, 2005, 2007 for bocaccio; 2005, 2007 for widow rockfish).

3. Previous STAR panel reports (2005, 2007 for bocaccio; 2005 for widow rockfish).

4. Stock Synthesis 3 (in a zip file, with a separate User Manual dated January 20 2009).

5. PFMC Terms of references for the groundfish stock assessment and review process for 2009–2010.

6. The following supporting papers and reports.

Field, J.C.; Ralston, S. (2005). Spatial variability in rockfish (Sebastes spp.) recruitment events in the California Current System. Can. J. Fish. Aquat. Sci. 62: 2199–2210

Hastie, J.; Ralston, S. (2007). Pre-Recruit Survey Workshop September 13-15, 2006 Southwest Fisheries Science Center Santa Cruz, California.

Ralston, S. (2009). Coastwide Pre-Recruit Indices from SWFSC and PWCC/NWFSC Midwater Trawl Surveys (2001-2008)

Ralston, S.; MacFarlane, B.R. (2009). Population estimation of bocaccio (*Sebastes paucispinis*) based on larval production

Ralston, S.; Pearson, D.E.; Field, J.C.; Key, M. (2009). Documentation of the California Catch Reconstruction Project.

APPENDIX 2: Statement of Work

This appendix contains the Statement of Work that formed part of the consulting agreement between Northern Taiga Ventures Inc. and the author.

Attachment A: Statement of Work for Chris Francis (NIWA)

External Independent Peer Review by the Center for Independent Experts

Stock Assessment Review Panel for Bocaccio and Widow Rockfish

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract to provide external expertise through the Center for Independent Experts (CIE) to conduct impartial and independent peer reviews of NMFS scientific projects. This Statement of Work (SoW) described herein was established by the NMFS Contracting Officer's Technical Representative (COTR) and CIE based on the peer review requirements submitted by NMFS Project Contact. CIE reviewers are selected by the CIE Coordination Team and Steering Committee to conduct the peer review of NMFS science with project specific Terms of Reference (ToRs). Each CIE reviewer shall produce a CIE independent peer review report with specific format and content requirements (**Annex 1**). This SoW describes the work tasks and deliverables of the CIE reviewers for conducting an independent peer review of the following NMFS project.

Project Description: Both bocaccio and widow rockfish have been declared overfished and are subject to rebuilding plans. The last benchmark assessment of bocaccio was conducted in 2003, using the Stock Synthesis 1 platform. The last benchmark assessment for widow rockfish was conducted in 2005, using a species-specific ADMB model. Both assessments will be conducted using the most recent version of the Stock Synthesis platform. These two benchmark stock assessments will provide the basis for the management of the groundfish fisheries off the West Coast of the U.S. including providing scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day meeting of fishery stock assessment experts. Participation of external, independent reviewer is an essential part of the review process. Participation of external, independent reviewer is an essential part of the review process.

The STAR panel is part of the Pacific Fishery Management Council's process to provide peer review as referenced in the 2006 Reauthorization of the Magnuson-Stevens Fishery Conservation and Management Act, which states that " the Secretary and each Regional Fishery Management Council may establish a peer review process for that Regional Fishery Management Council for scientific information used to advise the Regional Fishery Management Council about the conservation and management of the fishery (see Magnuson-Stevens Act section 302(g)(1)(E)). If a peer review process is established, it should investigate the technical merits of stock assessments and other scientific information used by the Council's Scientific and

Statistical Committee (SSC). The peer review process is not a substitute for the SSC and should work in conjunction with the SSC."

The Pacific Fishery Management Council's Terms of Reference for the West Coast Groundfish Stock Assessments and STAR Process for 2009-2010 requires that some reviewers be appointed from the Center for Independent Experts (CIE). The Council's terms of reference document will be included as background material. The Terms of Reference (ToRs) specific to the CIE are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Two CIE reviewers are required with one of the reviewers participating in all 2009 STAR panels (other than hake) to provide a level of consistency between the panels. The CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein. CIE reviewers shall have the expertise, background, and experience to complete an independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have expertise and work experience in fish population dynamics, with experience in the integrated analysis modeling approach, using age-and size-structured models, use of MCMC to develop confidence intervals, and use of Generalized Linear Models in stock assessment models.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Santa Cruz, California on July 13-17, 2009.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering committee, the CIE shall provide the CIE reviewer information (name, affiliation, and contact details) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and information concerning other pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., name, contact information, birth date, passport number, travel dates, and country of origin) to the NMFS Project Clearance for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer

review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations (available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send by electronic mail or make available at an FTP site the CIE reviewers all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewers shall read all documents in preparation for the peer review.

Documents to be provided to the CIE reviewers prior to the STAR Panel meeting include:

- The current draft stock assessment reports;
- The most recent previous bocaccio and widow rockfish stock assessments and STAR Panel reports;
- The Pacific Fishery Management Council's Scientific and Statistical Committee's Terms of Reference for Stock Assessments and STAR Panel Reviews;
- Stock Synthesis (SS) Documentation
- Additional supporting documents as available.
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

This list of pre-review documents may be updated up to two weeks before the peer review. Any delays in submission of pre-review documents for the CIE peer review will result in delays with the CIE peer review process, including a SoW modification to the schedule of milestones and deliverables. Furthermore, the CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein.

Panel Review Meeting: Each CIE reviewers shall conduct the independent peer review in accordance with the SoW and ToRs. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified in the contract SoW. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

In most circumstances a STAR Panel will include a chair appointed from the SSC's Groundfish Subcommittee and three other experienced stock assessment analysts. The STAR panel chair is responsible for: 1) developing an agenda for the STAR panel meeting, 2) ensuring that STAR panel members and STAT teams follow the Terms of Reference, 3) participating in the review of the assessment, 4) guiding the STAR panel and STAT team to mutually agreeable solutions, and 5) coordinating review of final assessment documents.

The CIE reviewer's role includes being an active panel participant and participants are strongly encouraged to voice all comments regarding the assessment data, model configurations, and uncertainty during the STAR Panel so the assessment teams can address the comments during the Panel meeting and incorporate changes when appropriate. The assessments are finalized by the end of the Panel meeting and comments made after the fact will not be able to be included in the final assessment document. The CIE reviewer should also contribute to the final STAR Panel Review Report. Additional details regarding the STAR Panel reviewer's responsibilities will be included in the Pacific Fishery Management Council's final Terms of Reference for Groundfish Stock Assessments and STAR Panel meetings.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer will assist the Chair of the panel review meeting with contributions to the Summary Report. CIE reviewers are not required to reach a consensus, and should instead provide a brief summary of their views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review;
2) Participate during the panel review meeting in Santa Cruz, California during July 13-17, 2009, as called for in the SoW, and conduct an independent peer review in accordance with the ToRs (Annex 2);
3) No later than July 31, 2009, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to David Die at ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in Annex 2;
4) CIE reviewers shall address changes as required by the CIE review in accordance with the schedule of milestones and deliverables.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

8 June 2009	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
29 June 2009	NMFS Project Contact sends the CIE Reviewers the pre-review documents
13-17 July 2009	Each reviewer participates and conducts an independent peer review during the panel review meeting in Santa Cruz, California.
31 July 2009	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
14 August 2009	CIE submits CIE independent peer review reports to the COTR
21 August 2009	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be made through the Contracting Officer's Technical Representative (COTR) who submits the modification for approval to the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the CIE within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and Terms of Reference (ToR) of the SoW as long as the role and ability of the CIE reviewers to complete the SoW deliverable in accordance with the ToRs and deliverable schedule are not adversely impacted. The SoW and ToRs cannot be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (the CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards: (1) each CIE report shall have the format and content in accordance with Annex 1, (2) each CIE report shall address each ToR as specified in Annex 2, (3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon notification of acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the approved CIE reports to the NMFS Project Contact and regional Center Director.

Key Personnel:

William Michaels, Contracting Officer's Technical Representative (COTR)
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

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

NMFS Project Contact:

Stacey Miller
NWFSC/FRAM Division
2032 SE OSU Drive, Newport OR 97365
Stacey.Miller@noaa.gov Phone: 206-437-5670

Elizabeth Clarke
NWFSC/FRAM Division
2725 Montlake Blvd. E, Seattle WA 98112
Elizabeth.Clarke@noaa.gov Phone: 206-860-5616

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a detailed summary of findings, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include as separate appendices as follows:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Stock Assessment Review Panel for Bocaccio and Widow Rockfish

1. *Become familiar with the draft bocaccio and widow rockfish stock assessments and background materials. Along with other members of the Panel,* determine if the stock assessment document is sufficiently complete according to the Pacific Fishery Management Council's Terms of Reference for West Coast Groundfish Stock Assessment and STAR Panels (to be included once finalized).
2. *Evaluate, data collection operations and survey design and make recommendations for improvement*
3. *Comment on quality of data used in the assessment.*
4. *Evaluate and comment on analytic methodologies*
5. *Evaluate model assumptions, estimates, and major sources of uncertainty. Specifically, recommend improvements including alternative model configurations or formulations as appropriate during the panel meeting and comment on the primary sources of uncertainty in the assessment model.*
6. *Insert an explicit statement as to whether this stock assessment represents the best available science.*
7. *Recommendations for any further improvements*
8. *Brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations*

Note – CIE reviewers typically address scientific subjects, hence ToRs usually do not involve CIE reviewers with regulatory and management issues unless this expertise is specifically requested in the SoW.

Annex 3: Tentative Agenda

Stock Assessment Review Panel for Bocaccio and Widow Rockfish
Santa Cruz, California

July 13-17, 2009

Point of contact for reviewer security & check-in: Dr. Steve Ralston

The agenda will be submitted to CIE at least two weeks before the meeting.

APPENDIX 3 Panel Membership

Participants in the review meeting included the following

Panel Reviewers

Martin Dorn, Scientific and Statistical Committee (SSC) Representative, Panel Chair

JJ Maguire, Center for Independent Experts (CIE)

Chris Francis, Center for Independent Experts (CIE)

Vladlena Gertseva, NMFS, Northwest Fisheries Science Center (NWFSC)

Panel Advisors

John DeVore, Pacific Fishery Management Council

Gerry Richter, Groundfish Advisory Subpanel (GAP) Representative

John Budrick, Groundfish Management Team (GMT) Representative

Stock Assessment (STAT) Team, Bocaccio

John Field, E.J. Dick, Alec MacCall. NMFS, Southwest Fisheries Science Center

Stock Assessment (STAT) Team, Widow rockfish

Xi He, E.J. Dick, John C. Field, Stephen Ralston, and Alec D. MacCall, NMFS, Southwest Fisheries Science Center

APPENDIX 4: Multinomial sample sizes

This Appendix (which was written for the STAR Panel report on widow rockfish) describes the method that was used to correct the multinomial sample sizes for compositional data in the Widow assessment, shows the results of its application to these data, and provides a brief rationale for the use of this method. The description that follows concerns age data, but the approach is exactly the same for lengths. Also, for simplicity, the following description ignores sex, but the extension to include sex is straightforward.

Suppose $p_{ay,\text{obs}}$ is the observed proportion at age a in year y in a set of age composition data that is assumed to have a multinomial error structure, and let $N_{\text{init},y}$ denote the initial sample sizes. We run our assessment model using these sample sizes and obtain, from the modwl fit, estimates of the expected proportions at age, $p_{ay,\text{exp}}$. Our aim is to use the $p_{ay,\text{obs}}$, $p_{ay,\text{exp}}$, and $N_{\text{init},y}$ to calculate a correction factor f so that the size of the model residuals is consistent with the corrected sample sizes, $N_{\text{corr},y} = fN_{\text{init},y}$.

This aim is the same as for the method currently used in SS3 to correct multinomial sample sizes. Where the two methods differ is that the SS3 method is based on the residuals for individual proportions, $r_{ay} = p_{ay,\text{obs}} - p_{ay,\text{exp}}$, whereas the present method is based on residuals for mean age, $r_y = (m_{y,\text{obs}} - m_{y,\text{exp}})$, where $m_{y,\text{obs}} = \Sigma_a(a\, p_{ay,\text{obs}})$, and $m_{y,\text{exp}} = \Sigma_a(a\, p_{ay,\text{exp}})$. The reason for using mean-age residuals is discussed below.

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

Figure A4.1 shows the application of this method to the Widow age and length composition data. In this application, the sample sizes, $N_{\text{init},y}$, were those obtained after correction (or tuning) using the SS3 method. That is to say, the residuals for individual proportions, r_{ay}, should be consistent with the size expected given these sample sizes. What Figure 1 shows is that the mean-age residuals are still too large (note that many of the confidence intervals for the observed values do not overlap the expected values). Thus, according to the present method, the sample sizes are too large, and so the estimated f is less than 1 for all data sets (range 0.04 to 0.45).

The reason the mean-age residuals, r_y, can be too large, while the individual proportion residuals, r_{ay}, are not, is that there is substantial correlation between the individual residuals. This is shown in Figure 2, in which the observed age frequency flips from one side of the expected frequency to the other from one year to the next. Sideways movement of this magnitude would not be possible if the individual residuals were uncorrelated. This correlation could be caused by either observation or process error (or a combination of both). One explanation of how this could occur derives from the fact that size (and thus age) distribution of fish often varies spatially. We would get between-year sideways movements of age frequencies if the spatial distribution of fishing changed substantially from year to year. In a model like that for Widow, in which selectivities are assumed to be time-invariant, this would movement would add process error to the observations. If the spatial distribution of catch sampling varied from year to year, in a way that did not reflect the movement of fishing activity, this would produce observation error, with a correlation between individual proportions. More information on the generation of correlations in composition data are given by Hrafnkelsson & Stefánsson (2004) and section 3 of Francis (2006).

Note that there is no intention to suggest that mean age (or length) is a quantity of particular interest in stock assessments. The only reason mean age (or length) is used in calculating the correction factor is that it is sensitive to the sort of correlations shown in Figures 1 and 2. Because of these correlations the composition data are less informative than is suggested by the multinomial sample size correction method used in SS3 (which assumes no correlation). Ideally, we should include these correlations in the likelihood function for composition data. However, that is not straightforward to do. The method proposed here is a simpler pragmatic alternative approach which adjusts sample sizes to compensate for the correlations.

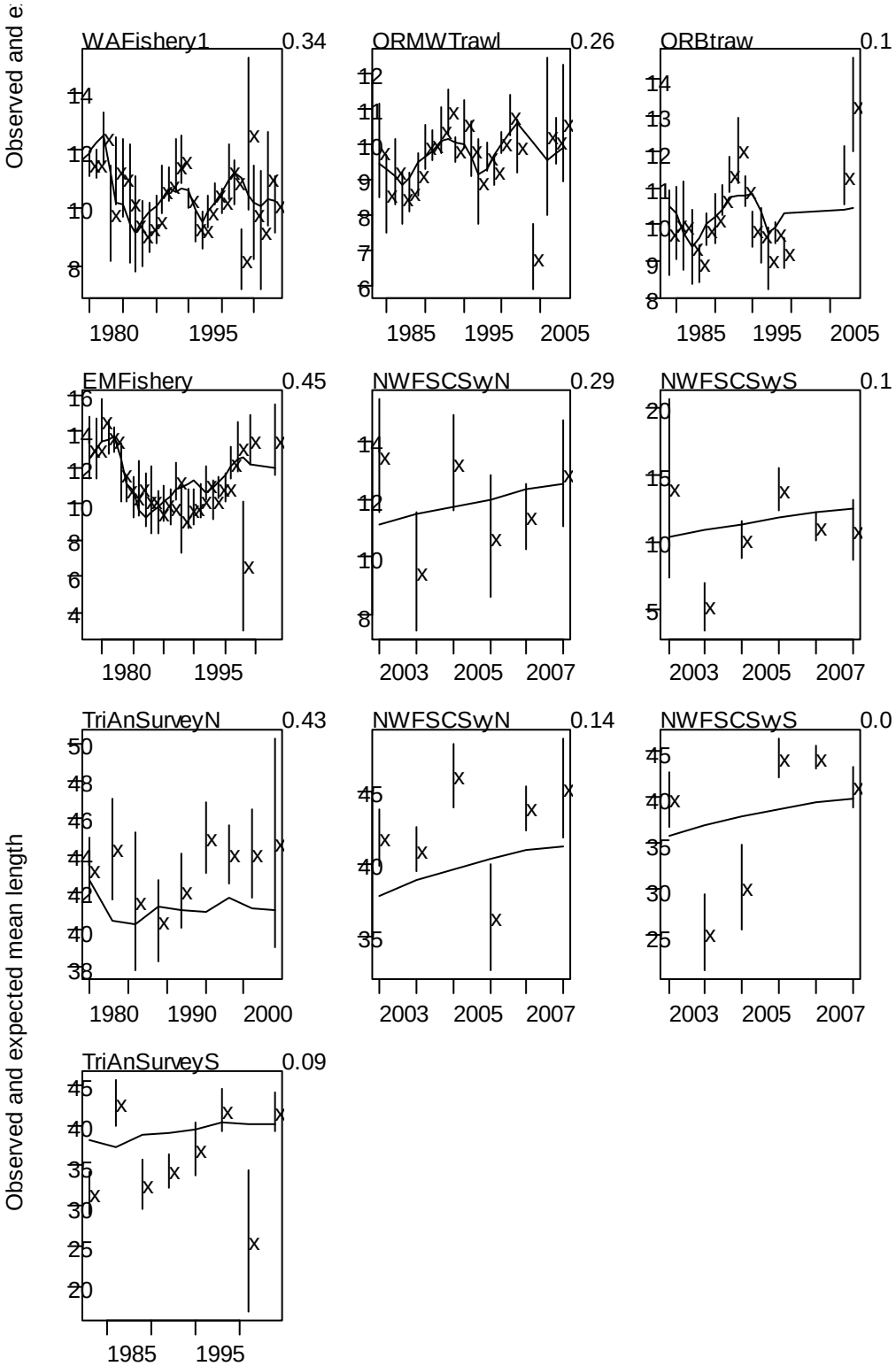

Figure A4.1: Observed ('x') & expected (line) mean ages (upper panels) and lengths (lower panel) for the Widow comps in the (former) base case model. Vertical bars are approximate 95% confidence intervals based on the multinomial sample sizes used in that model. The number printed above each panel is the correction factor *f*, calculated as described above.

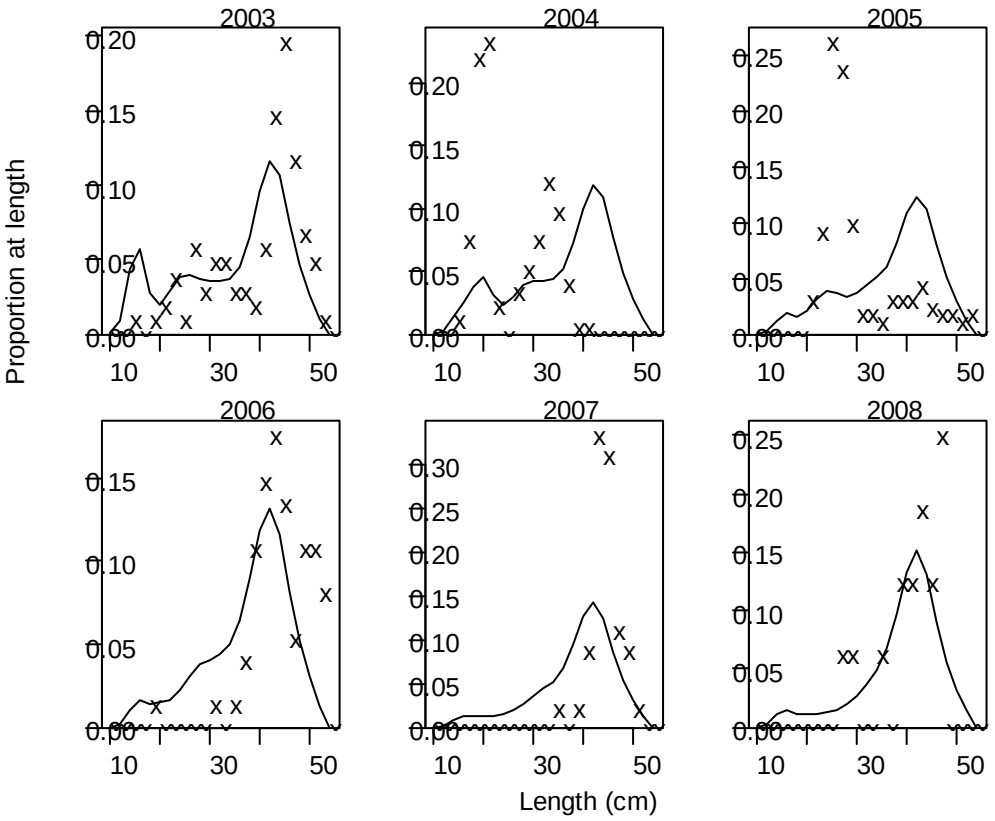

Figure A4.2: Observed ('x') and expected (line) proportions at length (sexes combined) for data set NWFSCSvyS.

References

Hrafnkelsson, B.; Stefánsson, G. (2004). A model for categorical length data from groundfish surveys. *Canadian Journal of Fisheries & Aquatic Sciences* 61: 1135-1142.

Francis, R.I.C.C. (2006). Some recent problems in New Zealand orange roughy assessments. *New Zealand Fisheries Assessment Report 2006/43.* 64 p.