Report of the Alaska Crab Stock Assessment Workshop[1]

EXECUTIVE SUMMARY

- A crab stock assessment workshop took place from 13-14 May 2009 at the Alaska Fisheries Science Center. Participation was by members of the Crab Plan Team (CPT) for the North Pacific Fishery Management Council (NPFMC), crab assessment authors, and other scientists involved in stock assessment and fishery management in Alaska.
- A set of guidelines was developed prior to the meeting and revised as needed during the workshop. These guidelines will form the basis for the May 2010 round of stock assessments, and will be refined from time-to-time to reflect the needs of the CPT and the NPFMC.
- A set of diagnostics and plots to assist the CPT in evaluating model fits was developed based on presentations by speakers with experience of stock assessments of species other than crab, as well as applications of candidate diagnostics to three representative crab stocks. These diagnostics are included in the set of guidelines for stock assessments.
- The original basis for the OFL control rule for Tier 4 stocks, $F_{MSY} \sim \gamma M$, was outlined and several alternative methods for determining γ for crab stocks were presented. None of these methods can be adopted at present, but example applications will be presented at the May 2010 CPT meeting.
- A series of recommendations were identified, defined as those to be implemented by the May 2010 CPT meeting and those which are longer-term.

A. INTRODUCTORY ITEMS

André Punt welcomed the participants (see Appendix A for a list of attendees) and outlined the Terms of Reference for the workshop:

(1) To standardize the crab stock assessments and assessment reporting to the extent possible given the inherent differences in the crab stocks and available data.

(2) To improve the crab stock assessments by resolving issues related to how data sources are weighted when an assessment includes several data sources (including the issues of diagnostics, residuals, and λ weighting).

(3) To determine how to calculate overfishing levels for Tier 4 stocks, including how to estimate γ, the natural mortality multiplier used to approximate F_{MSY}.

(4) To produce a workshop report that provides guidance to assessment authors to improve existing assessment models (snow crab, Bristol Bay red king crab, St. Matthew blue king crab, and Norton Sound red king crab) and to develop assessment models for stocks with sufficient data (Eastern Bering Sea Tanner crab, Aleutian Islands Golden king crab).

Punt noted that the workshop would only address items (1) – (3) and that a workshop report would be produced following the workshop. The report from the workshop, addressing item 4, will be presented at the September 2009 meeting of the CPT. The

[1] Prepared by André Punt and Doug Kinzey

draft agenda (Appendix B) was accepted without change, noting that the timing of agenda items would be modified as needed.

B. STOCK ASSESSMENT REPORTING

André Punt introduced a draft set of guidelines for the structuring of assessment documents. The draft was based on the groundfish terms of reference for stock assessments, similar documents used for Pacific Fishery Management Council stock assessments, and comments during past CPT meetings. The guidelines are aimed at authors producing reports reviewed by the CPT and other management bodies, and are intended to facilitate interpretation and review of assessments.

Workshop participants reviewed the draft set of guidelines, and several modifications were suggested. The final version of the document (Appendix C) was adopted intersessionally by CPT members.

The CPT noted that the guidelines do not indicate requirements for providing Acceptable Biological Catches (ABCs) and Annual Catch Limits (ACLs) in the crab stock assessment. The guidelines will need to be updated once these requirements are known. These guidelines are seen as a "living document" and will be updated as needed.

The workshop participants noted that stock assessment reports should include the model configuration (and associated data sources) on which management advice was based in the previous year. The status quo from the previous year provides a default that the CPT can select as the basis for management advice if updated or revised models appear to be unacceptable. Furthermore, the workshop **recommended** that the incremental affects of each change to a model and/or the data on which the assessment is based needed to be evaluated. This can be achieved by comparing results using the previously accepted model configuration against results after incorporating each new data source or model structural change. This process cannot be followed in all cases, for example, when a stock moves among Tiers, but every effort should be made to produce this type of diagnostic.

The workshop noted the importance of archiving of the software, data files, and assessment results so that, for example, historical retrospective analyses can be conducted. At present, this archiving is done by individual assessment authors, but there would be benefits if it was done at the agency level. In addition, software development would be enhanced if "version control" software was used – this helps to keep track of the changes made incrementally to assessment software, and is particularly useful when several assessment scientists are amending the same software.

Appendix C lists suggested model outputs and diagnostics statistics. Development of software which "automatically" produces the required plots and tables would make the process of model review (and report drafting) more effective. In this respect, the R routines developed by Dr Mike Prager and others (X2R; http://cran.r-project.org/contrib/extra/x2r/00ReadMe-X2R.html) which link ADMB output and R might provide a starting point for such software.

The discussions regarding stock assessment reporting focused on issues related to conducting stock assessments and calculating OFLs. However, ecosystem and economic information should also be included. It was noted that the extent to which ecosystem considerations are included in stock assessment reports was inconsistent among assessments, and that much of the information in these sections was outdated and perhaps even irrelevant for providing current management advice. The workshop recognized the need for, and the value of, a thorough evaluation of ecosystem considerations for crab stocks, but also recognized that there was been insufficient time for assessment authors to do the necessary work owing to the substantial model development work in recent years.

The workshop consequently **recommended** that a two step process be undertaken to improve the analyses of ecosystem considerations:

(1) Development of a separate, general document which outlines the issues (such as the impact of climate change, physical factors, etc.) affecting all crab stocks and the general impacts of crab fisheries on other aspects of the ecosystem.
(2) Development of specific ecosystem considerations chapters for each species, highlighting information and issues specific to each crab stock.

The workshop **recommended** that the overview document be developed before the May 2010 CPT meeting, with the May 2010 meeting including time to review this document. Bob Foy indicated the NOAA Kodiak Lab would take the lead in developing the ecosystem document.

C. DATA WEIGHTING AND DIAGNOSTICS

André Punt noted that stock assessments results depend on how the data sources are weighted. Consideration of the relative weighting of different data sources is most important when the data appear to be in conflict (although in that case, it is also important to show results for subsets of the data which are consistent). Data weighting also impacts the perceived precision of the outcomes from assessments.

Data overweighting systematically improves estimate perceived precision, and will consequently bias selection of preferred methods. In general, using the raw sample sizes (e.g., number of crabs measured) for size or age composition data, and sample standard deviations for indices of abundance will overweight the data because: (a) non-random selection of sampling units and clustering of the population will underestimate uncertainty if raw sample sizes are used for data weight, and (b) systematic effects are not represented in sampling standard deviations (e.g. serially-correlated temperature impacts on survey catch rates).

André noted that the ideal diagnostic statistics and plots for evaluating data weighting should:

(1) allow inconsistencies between data and model predictions to be identified;
(2) be simple to view and understand (particularly by non-modellers);
(3) be automatically produced by the assessment software (so that diagnostics can be produced “on the fly” during assessment reviews); and
(4) allow ready identification of overdisperson (in the observation error variances for the abundance indices and the effective sample sizes for the compositional data) and whether assumptions on recruitment variability are supported by the assessment results.

C.1 Presentation summaries

C.1.1 Pete Hulson

Data weighting corresponds to the uncertainty in observations that are fit in stock assessment models to estimate population parameters. An intuitive interpretation of data weighting is that the weighting term used in the objective function is inversely proportional to the dataset uncertainty. A general formulation of the objective function, *O*, is:

$$O = \sum_{X} \lambda_x G\left(Y_x, \hat{Y}_x\right) \tag{1}$$

where λ_x is the weighting term and $G(Y_x,\hat{Y}_x)$ is some function that relates the observations, Y_x, to the model predictions, $\hat{Y}_x$. The combination of the weighting term and function *G* is defined as a likelihood function for some assumed statistical distribution, often evaluated as the negative log-likelihood.

Two primary structures of data are fitted in stock assessment models: (1) index data (e.g., survey biomass, catch-per-unit-effort, total fishery catch); and (2) compositional data (e.g., catch-at-age and/or catch-at-length, survey proportions-at-age/length). A lognormal distribution is often (but not always) used to fit index data, and after omitting constants the likelihood is given by:

$$\lambda_x G\left(Y_x, \hat{Y}_x\right) \cong \frac{1}{\sigma^2_{Y_x}}\left(Y_x - \hat{Y}_x\right)^2 \tag{2}$$

In this case, the weighting term λ_x is interpreted as the inverse of variance of the observed index data. A multinomial likelihood is often used for compositional data, and can be evaluated as:

$$\lambda_X G\left(Y_X, \hat{Y}_X\right) \cong n_X \sum_{a} Y_{a,x} \ln \hat{Y}_{a,x} \tag{3}$$

where the weighting term is interpreted as the sample size, n_X .

While these are theoretical definitions for the dataset weighting, uncertainty in observations is usually unknown for fishery data. Standard practice is for the stock assessment scientist to pre-specify the weighting term used in the objective function. The literature does provide some more objective methods to determine weighting. For index data, the weighting can be determined by:

(1) sampling uncertainty (Sullivan *et al.*, 1999);
(2) expert opinion (Merritt and Quinn, 2000); or
(3) model estimation (Kimura, 1989; Maunder and Starr, 2003)

For compositional data, an effective sample size replaces the sample size as the weighting. The effective sample size is usually smaller than the actual sample size due to violations of multinomial processes that cause over-dispersion of the data. Some methods used to define effective sample sizes for multinomially-distributed data include:

(1) setting all effective sample sizes to 400 (Fournier and Archibald, 1982);
(2) sample size, capped at 1000 (Fournier *et al.*, 1998);
(3) setting all effective sample sizes to 200 (Methot, 2000);
(4) setting the annual effective sample size to the square root of annual sample size (Hanselman *et al.*, 2007);

(5) estimating the effective sample size within the model as a parameter (Fournier *et al*., 1990);
(6) using iterative estimation (McAllister and Ianelli, 1997);
(7) estimating the effective sample size based on sampling uncertainty (Crone and Sampson, 1998); and
(8) estimating the effective sample size based on the Dirichlet distribution (Williams and Quinn, 1998).

C.1.2 Jim Ianelli
Jim Ianelli (AFSC) discussed data influences on statistical models of different complexity and presented examples of "historical" retrospective analyses. Jim noted that he has included "traditional" retrospective analyses (applying the same model to reduced datasets) in previous assessments. The presentation emphasized two scenarios. The first scenario involved a simulated two-index model where trend is the parameter of interest. This two-index model simultaneously analyzed a "noisy" declining index and a relatively precise index with no trend. Given equal weights and estimating both the variance terms and the trends resulted in a fit that effectively ignored the noisy index (very similar results occurred when the variances were pre-specified). When the weights were changed (and variances estimated), a bimodal likelihood profile caused the model to focus on the heavier weighted index and ignore the second index. This bimodality disappeared when variances were pre-specified. In conclusion: (a) variances should be estimated (implicitly or explicitly) with caution; (b) residuals should always be examined; and (c) model specifications should be examined.

A second scenario examined the influence of new data on determination of biological reference points (e.g., the ABC) for the Eastern Bering Sea pollock stock. In 2008, the ABC was quite sensitive to new data because the stock was below the target level and the harvest control rule ratchets down acceptable fishing mortality rates. Options for fitting models to different data sources were presented.

As an alternative for evaluating "effective N" for compositional data, the observed mean age (or length) for a given year and gear can be plotted with implied confidence bounds using:

$$Var\left(\bar{X}\right)=\frac{\sigma^2}{n} \qquad (4)$$

(where σ is the standard deviation and n is the sample size) and then compared with model predicted mean age (or length).

C.1.3 Martin Dorn
The assessment for Gulf of Alaska pollock is an integrated assessment that uses trend data from multiple surveys, and age composition data from the fishery and from fishery-independent surveys. Given that data come from different sources, not all data sets are equally informative about stock status. An initial step in any assessment is a careful look at the data before fitting models, with attention given to identifying contradictory data. The pollock assessment involves an age-structured population model using maximum likelihood estimation to fit available data. Arbitrary data weighting terms (λs) are not used for likelihood components; instead more specific likelihood-related terms are used. For trend data modelled with a log-normal

likelihood, the survey CV (or log standard error) is the basic measure of uncertainty. For compositional data modelled with multinomial likelihoods, the input sample size is the basic measure of uncertainty. This approach makes it possible to evaluate goodness of fit using standard summary statistics. The root mean squared error, $RMSE = \sqrt{\frac{1}{n}\sum(\ln(obs) - \ln(pred))^2}$, where n is the number of data points, is used to summarize the fit to a survey time series. For compositional data, the effective sample size, $effN = \frac{\sum p\,(1-p)}{\sum(obs-pred)^2}$, where p is the proportion, is often used. For model tuning the input survey CVs and multinomial sample sizes are adjusted to be comparable to the summary statistics of model output. Any tuning should preserve the differences in uncertainty within data sets, such as annual survey estimates with unusually high CVs, or age-composition samples with low sample sizes. A pragmatic approach for model tuning should make input and output statistics commensurate, but also acknowledge that good reasons may exist for accepting some lack of model fit, for instance when deciding to use contradictory data sets.

C.1.4 Cathy Dichmont

Two case studies from Australia were provided for discussion. The example, of the Northern Prawn Fishery (NPF) highlighted different diagnostics and how likelihoods can be set up to include, amongst others, effective sample sizes. The NPF is a multi-species tropical fishery targeting short-lived shrimp species. The two “fleets” in this fishery catch a group of overlapping species. Management is based on a target reference point intended to maximise discounted profit over a 25-year period. The limit reference point is biologically based. Available data include catch and effort by species since 1970, catch size-composition data from observers, biannual survey indices of abundance and associated size-composition data (since 2002), and economic survey data. The size-structured population dynamics model operates on a weekly time step. The catch likelihood assumes that the square root of the observed catch is normally distributed. The survey index includes two variance components: sampling and “other” error. This is because the observed survey variance is smaller than the true variance. The size-composition likelihood assumes that the length-frequency data are multinomially distributed about the model predictions. It also accounts for an effective sample size parameter, which is a corrected sample size (smaller than the raw sample size) that takes the fact that size measurements are correlated (e.g. through schooling behaviour) and therefore the length-frequency data provide less information than if the samples were randomly selected from the population (see Folmer and Pennington (2000)). Tagging data are used to estimate growth (Punt *et al.*, 2009). Growth from size class i to j is assumed to follow a normal distribution, but one where the variance parameter for sizes less than the size at maximum selection may differ from that for sizes greater than the size at maximum selection. In this example there are many residuals to look at - multiple data sources, weekly time step, 3 species, 2 fleets, 2 sexes. Example diagnostics were shown, such as summarising over some of the components e.g. year, week.

The second example, school whiting in the South East Scalefish and Shark Fishery (SESSF), demonstrates the impacts of different data sources on changes to the likelihood weightings. The SESSF is a multi-species and multi-fleet fishery. The school whiting assessment is based on Stock Synthesis 2 (SS2). Two state fleets

(Danish Seine and otter trawl) fleet fish this resource. Data sources include logbook catch and effort by fleet, onboard observer length-frequency (discard rate) data, fish market length-frequency data (1983-89), port-based length-frequency data, standardised catch rates for the Danish seine fleet, age-at-length data, age-frequency data, information on age reading error, and discard rates (2004-present). This age- and length-structured assessment uses the iterative weighting method available in SS2: the CV of Danish seine fleet catch rates is iteratively adjusted so that the observed CV is the same as the expected CV, and the length and age data weightings are set using iterative reweighting of these data (to match input and output effective sample sizes). A plot of effective and observed sample size was shown as a diagnostic test. Extensive sensitivity tests of the management outputs, such as present stock status and recommended biological catch, are routinely conducted. These tests include halving and doubling the weighting on catch rates, length frequency data and age composition data to examine the sensitivity of the results to these data sources.

C.1.5 André Punt

André noted that the CPT has had difficulty interpreting "bubble plots" (plots of standardized residuals based on fits to the size-composition data from surveys and the catch). He therefore introduced a potential algorithm (based on an approach outlined in Peacock (1983)) as a tool for evaluating whether such plots indicate "random" residuals:

(1) Normalize the residuals so that they have mean 0 and variance 1 (whether residuals do or do not have mean 0 and variance 1 should be identified using another test – this test merely addresses the randomness issue).

(2) Denote the residual for year y and size-class k as $r_{y,k}$ and compute the quantity:

$$\tilde{r}_{y,k} = r_{y,k} - \tfrac{1}{8}[\sum_{y'=y-1}^{y+1} \sum_{k'=k-1}^{k+1} r_{y',k'} - r_{y,k}] \tag{5}$$

(3) Compute a cumulative distribution for $\tilde{r}_{y,k}$ based on the observed data and for $\tilde{r}_{y,k}$ had the $r_{y,k}$ been iid N(0,1) random variables.

(4) Plot the two cumulative distributions.

André evaluated this method for a few example patterns (See Appendix D). The meeting welcomed the approach and encouraged additional analyses to further evaluate it.

C.2. Exploratory comparisons

Three of the assessment authors (Jie Zheng, Shareef Siddeek, and Jack Turnock) were requested to conduct analyses for Bristol Bay red king crab (BBRKC), Aleutian Islands golden king crab (AIGKC) [east of 174^0W only], and EBS snow crab, respectively, to explore the value (and implications) of different diagnostic statistics. The workshop agreed that the analyses were for illustrative purposes only and would not impact the decisions made regarding models by the CPT, because there was insufficient time to evaluate any revised model formulations. The requests were:

(1) Provide a list of weights assigned to the indices and compositional data, in the form of standard deviations for the indices and the number of independent samples for the compositional data, and list any other weights in the assessment.

(2) Compute the "effective" sample sizes for the compositional data using the formula:

$$n_y = \sum_l \hat{P}_{y,l}(1-\hat{P}_{y,l}) / \sum_l (P_{y,l} - \hat{P}_{y,l})^2 \qquad (6)$$

where n_y is "implied" effective sample size for year *y*;

$P_{y,l}$ is the observed proportion of the catch of animals in length-class *l* during year *y*; and

$\hat{P}_{y,l}$ is the model-estimate of the proportion of the catch of animals in length-class *l* during year *y*.

Plot the assumed ("input") and "effective" sample sizes

(3) Compute the Root Mean Square Errors (RMSEs) for the fits to the indices using the formula:

$$\sigma = \sqrt{\tfrac{1}{n}\sum_y \left(\ln I_y - \ln \hat{I}_y\right)^2} \qquad (7)$$

where I_y is the observed index for year *y*;

$\hat{I}_y$ is the model-estimate corresponding to I_y, and

n is number of data points for index.

(4) Plot the time-trajectory for mature male biomass (at the time of mating) when the weight assigned to each data source (and the weight on the penalty on the extent of inter-annual variation in recruitment) is doubled.

(5) Plot the marginal (over year) observed and predicted distributions of catch (or survey)- proportions at length.

It was not possible for the assessment authors to conduct all of the requested analyses in the time available (overnight), but there were sufficient results (see Appendix E) for the workshop to be able to draw some key conclusions:

- The root mean square errors, RMSEs, about the survey indices were markedly larger than the pre-specified coefficients of variation for these data for EBS snow crab and BBRKC. For example, the RMSE was 0.303 for BBRKC while the pre-specified CV was only 0.2. Similarly, the RMSE for the snow crab survey indices markedly exceeded the pre-specified CVs for these indices.
- There were a few instances where it appears that the implied effective sample sizes for the size-composition data were notably different from the assumed values (for example, the retained catch of AIGKC and the retained catch of BBRKC).
- The results for snow crab were not particularly sensitive to changing the weights assigned to the data. In contrast, varying some of the data weights in the AIGKC assessment had a marked impact on the results.
- There is value in plotting the time-series of implied effective sample sizes, in addition to plotting these using a histogram or as plots of implied versus input effective sample sizes.
- The marginal distributions of catch (and survey) proportions exhibit systematic patterns of deviation for all of these assessments (although not for

all sources of data in each assessment, and the extent to which there is evidence for mis-specification differed among assessments).

C.3. Discussion

All crab assessment models include "penalties" (i.e. constraints on the values for parameters such as annual recruitment and fishing mortality). In discussion, it was noted that this would invalidate both methods for estimating variance using asymptotic methods and approaches for comparing models (including likelihood profile and AIC). It was also noted that it is not valid to compare model fits and outputs for two assessments which use a different mix of data sets.

Jim Ianelli noted the value of comparing the pre-specified value for the variance in recruitment (σ_R) with the variance of the estimates of recruitment from the assessment. He further noted that the variance of the estimates of recruitment from the assessment model will under-estimate the recruitment variance when there is a penalty on recruitment in situations where there is little information on year-class strength

The workshop noted that a variety of ways have been employed for specifying input sample sizes for compositional data in crab assessments, although several other methods exist. The meeting characterized the methods as follows:

(1) a fixed constant (dependent on data-type);
(2) based on bootstrapping using the design of the sampling scheme;
(3) number of hauls, tows, or trips (perhaps approximated by dividing the number of animals sampled by a constant);
(4) as for (3), except that a maximum sample size is also imposed; and
(5) the number of animals sampled divided by the maximum sample size, and multiplied by a pre-specified constant.

The meeting discussed different ways of adjusting the input sample sizes.

(1) Martin Dorn noted that he did not tune his effective sample sizes, but rather examined the input and implied effective sample sizes to check that they are in the same "ballpark". He noted further that he would not necessarily adjust the input CVs for indices even if they were markedly different from the implied CV. For example, two of the indices used in the GOA pollock assessment were in conflict during the early years of the assessment period, which leads to large residuals in both indices, and large implied CVs.
(2) Jim Ianelli noted that he preferred to set weights before applying the model because any "tuning" algorithm relies on the assumption that the population and observation model are correct.
(3) Jim Ianelli noted that when "tuning" the CVs assumed for the indices, it is best to fit the model setting the CVs to pre-specified values during the early phases of the estimation and only adjust these sizes in the final phases of the analyses.

The workshop noted that the multinomial likelihood is not robust to outliers and outliers may therefore have an important impact on the results from stock assessments and hence on the selection of weighting schemes. It **agreed** that there was value in exploring the impact of assuming different likelihood functions, and in particular formulations which should be more robust to outlying observations. Other alternative likelihood functions for index and catch data include assuming that the square roots of the data are normally distributed (e.g. Dichmont *et al.*, 2003) and that the data are approximately chi-square distributed. In principle, the effective sample size for a multinomial distribution can be estimated using maximum likelihood, but this requires including all of the "constants" when coding the likelihood function. In this respect, an alternative to the multinomial likelihood would be the Dirichlet distribution (Williams and Quinn, 1998) or the robust likelihood function of Fournier *et al.* (1998).

The workshop discussed how to deal with cases when diagnostics such as those in Section C.2 indicate that assumptions appear to be violated. Two main approaches emerged (although reality will lie between these two philosophies, and both approaches have been applied when conducting assessments of BSAI crab stocks):

(1) The data (or their weighting) are wrong; the solution in this case is to change (generally reduce) the weights assigned to the data (tune the effective sample sizes and survey CVs) until the diagnostics show no problem or, if some of the data sets are in conflict, to present assessments based on subsets of the data which are not in conflict.
(2) The model is wrong; the solution in this case is to change the model (generally allow for more flexibility, such as more time blocks for selectivity and growth) or the likelihood function.

Irrespective of which of these two approaches is taken, the aim should be that the final model is "roughly" consistent in terms of the diagnostics listed above. However, there will be reasons why the data may not be fully consistent with the model in an acceptable assessment (e.g. GOA pollock).

The meeting **agreed** that whenever possible:

(1) weights should be expressed as standard deviations or effective sample sizes to ensure comparability among assessments (these should reflect both the variation in sampling and the validity of the assumptions of the model to the extent possible); and
(2) weights should not be set higher than implied by the extent of sampling error (e.g. by setting the CVs for survey indices lower than the CVs inferred from the data collected from the survey).

The meeting made the following additional **recommendations**:

(1) The stock assessment guidelines should be modified to include the types of diagnostic statistics considered above (see Section E.4.4 of Appendix C).
(2) Sensitivities to weights should be conducted whenever a model is modified, but there is no need to examine this sensitivity very often for "fully developed" models.
(3) André should work with the assessment authors to specify specific scenarios to consider when examining sensitivity to weights.
(4) Consider developing the facility to estimate the extent of "additional variance" for the survey indices (the difference between the assumed and implied CVs) within assessments.
(5) Compare the input and implied values for the extent of variation in recruitment, σ_R.

D. OVERFISHING LEVELS FOR TIER 4 STOCKS

Overfishing is defined as any amount of catch in excess of the OFL. Overfishing for BSAI crab is determined by comparing the OFL, as calculated in the five-Tier system (Table 1), for a crab fishing year with the catch estimated for that year.

The Tier 4 OFL control rule is for stocks where essential life-history and recruitment information, and understanding, are lacking. There is information about basic life-history parameters (e.g. growth, natural mortality, and maturation) and an index of abundance (typically from an assessment model), but no stock-recruitment relationship for stocks in Tier 3, while this information in addition to a reliable stock-recruitment relationship are available for stocks in Tiers 1 and 2 (there are no such BSAI crab stocks at present). There are no reliable estimates of biomass or *M* for Tier 5 stocks. Table 2 lists the ten BSAI crab stocks by Tier level, five of which are currently assigned to Tier 4.

Explicit to Tier 4 are reliable estimates of biomass (either from surveys or an assessment model) and the instantaneous rate of natural mortality, *M*. The proxy for B_{MSY} for Tier 4 stocks is the average biomass of mature males at the time of mating over a specified period. The OFL control rule for Tier 4 stocks involves multiplying *M* by a parameter, γ, to estimate the OFL fishing mortality, F_{OFL}. γ is allowed to be less than or greater than unity. Use of γ is intended to allow "adjustments in the overfishing definitions to account for difference in biomass measures" (Anon, 2008), but also accounts for, for example, differences between the maturity and selectivity patterns. The final rule implementing the revised OFL harvest control rule set the default value for γ at 1, with the understanding that the Council's SSC may recommend a different value for a specific stock or stock complex, as merited by the best available scientific information.

Among the purposes of the workshop was the explicit aim to explore methods for assigning an appropriate value for γ for Tier 4 stocks. Several participants noted that the default for γ of 1 was included in the EA at the NMFS review stage with no supporting analysis. Moreover, values for γ for modelled stocks evaluated in the OFL EA exceeded 2, primarily because of the growth dynamics of crab – a terminal molt for some species and the differences between male maturation and selection to the fishery. Although basing stock status determination on abundance of mature male biomass was viewed as conservative; it was noted that some crab stocks had declined substantially under what appears to have been fairly low levels of fishing mortality. Consequently, appropriate values for Tier 4 stock γ levels remain unresolved.

D.1. Presentation by Terry Quinn

The natural mortality parameter *M* has been used as a proxy for F_{MSY}, dating back to at least the 1960's (e.g., Alverson and Pererya, 1969; Gulland, 1970). However, Deriso (1982) showed that F_{MSY} is less than *M* for many parameter combinations in a delay-difference model. Thompson (1993) showed that fishing mortality should be less than 80% of *M* for spawning biomass per recruit to remain above 30% of the pristine level, one benchmark used as an indicator of an overfished population. Thus, *M* should not be a target but may work as a limit. And this is exactly how *M* is treated in the Tier system for both groundfish (Tier 5) and crab (Tier 4).

Given the complicated life history of crabs and the unusual male-only fishery, it is not known just how far fishing mortality should be set away from M. Therefore, a special coefficient, γ, was included to allow adjustments due to different biomass measures, size limits, harvest strategies, and other potential factors. Thus the OFL fishing mortality for Tier 4 was written as $F_{\text{OFL}} = \gamma M$. A crab workshop was held in February/March 2007; the workshop report clarified: "In the new Tier 4 (previously Tier 5), a scalar γ is multiplied by natural mortality. The scalar could be less than or greater than 1 and be more or less conservative than the status quo, depending on stock assessment research for a species. For example, when a change from total mature biomass to some other biomass measure (e.g., based on mature males) is used, the scalar can be applied to account for differences between biomass measures."

Analysis in the Environment Assessment for Amendment 24 (the November 2007 version) to revise overfishing definitions for crab showed that values of γ between 2 and 3 were appropriate for F_{MSY}. There was no default value for γ in the EA reviewed by the Council family in October 2007. However, a default value of 1 was inserted into the Final Environmental Assessment (dated May 2008). It is not clear to me who was involved in deciding on the default value. I could find no discussion of this in SSC emails or reports.

The June 2008 SSC report discussed the preliminary 2008 BSAI crab SAFE produced by the Crab Plan Team, discussed that γ could be less than or greater than 1, and called for more quantitative analysis in future years. The Crab Plan Team produced the revised BSAI crab SAFE in September 2008, in which actual calculations of OFL were made. Of the 10 stocks listed in Table 3 of that document, five were in Tier 4, and the values of γ selected for those in stock status levels a and b (four stocks) were all equal to 1. The rationale for each stock was given in each SAFE chapter, but the common theme was to be conservative in the face of data and population uncertainties. And thus, the SSC has gone along with this approach until better approaches are available.

D.2. Presentation by André Punt

André Punt noted that an objective method was needed to specify an F_{MSY} proxy for use when applying the Tier 4 control rule and that there was also a need to calculate a B_{MSY} proxy for Tier 4 (and 3) stocks. Both of these tasks had been difficult for the CPT in the past. André introduced two methods for calculating F_{MSY}. One of these is based on information about selectivity, growth, maturity and natural mortality and has been applied to data for Tanner crab (Turnock and Rugolo, 2008). The second method estimates F_{MSY} proxies based on survey data and assessment output (Appendix F). André provided an initial analysis of the ability of the latter approach to correctly estimate F_{MSY} (Appendix F). The performance of this second method has not yet been fully evaluated, but it appears that its results are fairly sensitive to the extent of both observation and process error.

André also mentioned that it should be possible to estimate the proxy for B_{MSY} for Tier 3 and 4 stocks (B_{REF}) by projecting the assessment model forward with recruitment selected at random from the years used to define B_{REF} and fishing mortality set equal to the proxy for F_{MSY}. While not ideal because it fails to account for any stock-recruitment effect, this approach determines a B_{MSY} value that is consistent with the F_{MSY} proxy and the recruitments used to define B_{REF} without being influenced by the historical harvest (which may have been zero or unsustainable).

D.3. Discussion

The workshop **agreed** that the values for γ in the EA can only be assumed to apply to stocks other than those for which they were derived if growth, selection and maturity were similar for the stock for which the estimate of γ was derived and the stock for which a value for γ was needed. This is difficult to show in general. However, Amendment 24 to Crab FMP does allow alternative methods for specifying γ, if these can be justified.

In relation to the approach in Appendix F, the meeting noted that the approach is empirical and would require a definition for "male mature biomass". For stocks for which assessments are available, this would be mature male biomass as defined in the assessment while ogives are available to define mature male biomass for surveys. The meeting noted this approach had promise and **recommended** that assessment authors should attempt to apply it. However, it needs to be more fully evaluated, including whether weighting different data points might lead to improved performance (as would be expected if there were no observation and process error).

In relation to the Turnock and Rugolo (2008) method, the meeting noted that it should be possible to compute survey selectivity and male maturity-at-length for most stocks (at least roughly) and proxies for M exist for most stocks. However, it may be more difficult to specify growth for several of the Tier 4 stocks. Nevertheless, the meeting **recommended** that assessment authors should attempt to apply this method (see Appendix G for details). As for the method in Appendix F, it is necessary to more fully evaluate this approach. Ultimately, this approach could be applied to all Tier 4 stocks.

Thompson (1992) provides an approach for estimating F_{MSY} using only an assumption about the compensation of the stock-recruitment relationship, r, the rate of natural mortality, M, and the difference between the age at maturity and the age intercept of the linear weight-at-age equation, d. Anon (2009) used this approach to estimate F_{MSY} for snow crab in the Arctic.

In the longer-term, the meeting **recommended** that it would be useful to further investigate the likely range for $F_{35\%}/M$ and F_{MSY}/M for crab stocks. Two approaches identified for this were:

(1) construct generic models for crab stocks and explore how the values for these ratios change as a function of different biological (and fishery) assumptions. The results of this study could be used to assign crab stocks likely ranges for γ based on how they are categorized; and
(2) conduct a meta-analysis of F_{MSY}/M for crab fisheries worldwide.

E. SUMMARY OF RECOMMENDATIONS

E.1. Short-term (by May 2010)

1. Assessments should be based (to the extent possible) on the guidelines in Appendix C.
2. Include the following diagnostics in the stock assessments:
    1. marginal distributions of fits to compositional data;

   2. plots of implied and assumed effective sample sizes for compositional data;
   3. tables of RMSE values for the abundance indices;
   4. q-q plots, histograms of residuals for the compositional data and abundance indices (separately by data series); and
   5. results when the weights for each data input are systematically increased / decreased.
3. Avoid the use of λs (emphasis factors) to the maximum extent possible and instead report the weights in the form of CVs or effective Ns.
4. Document the basis for the input effective sample sizes.

E.2 Medium-term (ideally by May 2010)

1. Develop the facility to estimate the additional variation of surveys (add variance to CVs) within assessments (i.e. automatic retuning).
2. Compare input and output values for σ_R.
3. Consider alternative error distributions [e.g. adding small constants / the robust normal likelihood for proportions].
4. Apply the Appendix F and G approaches to all Tier 3 and 4 stocks.
5. Review the ability of the ratio of fishery catches to survey catches (by length) as a way to estimate selection curves (Tier 4 stocks).
6. Compute B_{REF} by projecting models forward under F_{MSY} (all assessed stocks).

E.3 Longer-term

1. Consider the establishment of an assessments methods working group, with participation from assessment scientists working on West Coast and Alaskan stock assessments.
2. Further develop methods for assessing the randomness of bubble plots.
3. Explore the use of the Dirichlet likelihood function.
4. Develop methods which can automatically produce the types of diagnostic statistics listed in Appendix C (e.g. based on an R package).
5. Further simulation testing of the method of Appendix F and the Turnock and Rugolo (2008) approach.
6. Further generic evaluation of $F_{35\%}/M$ and F_{MSY}/M for crab-related life histories.
7. Conduct a meta-analysis of F_{MSY}/M for crab fisheries worldwide.
8. Evaluate the utility of the approach of Thompson (1992) for crab stocks.

F. CLOSING REMARKS

André Punt thanked the participants, especially the assessment authors (Jie, Jack and Siddeek) who conducted overnight analyses and the invited speakers, for what was a very productive workshop, the outcomes of which should help the CPT with its work. He noted that the workshop report would be finalized in the next few months and presented to the September 2009 meeting of the CPT.

G. REFERENCES

Alverson, D.L. and W.T. Pereyra. 1969. Demersal fish explorations in the northeastern Pacific Ocean - An evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. *Journal of the Fisheries Research Board of Canada* 26: 1985-2001.

Anon. 2009. Fishery Management Plan for Fish Resources of the Arctic Management Area. North Pacific Fishery Management Council, 605 W. 4th Ave, Suite 306, Anchorage, Alaska, 99501.

Crone, P.R. and D.B. Sampson. 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. pp. 355-370. In: Fishery stock assessment models. F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C. I. Zhang (eds), Alaska Sea Grant Program Report No. AK-SG-98-01, University of Alaska Fairbanks

Deriso, R.B. 1982. Relationship of fishing mortality to natural mortality and growth at the level of maximum sustainable yield. *Canadian Journal of Fisheries and Aquatic Sciences* 39: 1054-1058.

Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q. and W. Venables. 2003. Application of a weekly delay-difference model to commercial catch and effort data in Australia's Northern Prawn Fishery. *Fisheries Research* 65: 335-350.

Folmer, O. and M. Pennington. 2000. A statistical evaluation of the design and precision of the shrimp trawl survey off west Greenland. *Fisheries Research* 49: 165-178.

Fournier, D. and C.P. Archibald. 1982. A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences* 39: 1195-1207.

Fournier, D.A., Hampton, J. and J.R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunge*. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 2105-2116.

Fournier, D.A., Sibert, J.R., Majkowski, J. and J. Hampton. 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency datasets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). *Canadian Journal of Fisheries and Aquatic Sciences* 47: 310-317.

Gulland, J.A. 1970. The fish resources of the ocean. FAO Fisheries Technical Paper 97.

Hanselman, D.H., Heifetz, J., Fujioka, J., Shotwell, S.K. and J.N. Ianelli. 2007. Gulf of Alaska Pacific ocean perch. In Stock Assessment and Fishery Evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2008. North Pacific Management Council.

Hilborn, R. 2001. Calculation of biomass trend, exploitation rate, and surplus production from survey and catch data. *Canadian Journal of Fisheries and Aquatic Sciences* 58: 579-584.

Kimura, D.K. 1989. Variability, tuning, and simulation for the Doubleday-Deriso catch-at-age-model. *Canadian Journal of Fisheries and Aquatic Sciences* 46: 941-949.

Maunder, M.N. and P.J. Starr. 2003. Fitting fisheries models to standardized CPUE abundance indices. *Fisheries Research* 63: 43-50.

McAllister, M.K. and J.N. Ianelli. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* 54: 284-300.

Merritt, M.F. and T.J. Quinn II. 2000. Using perceptions of data accuracy and empirical weighting of information: assessment of a recreational fish population. *Canadian Journal of Fisheries and Aquatic Sciences* 57: 1459-1469.

Methot, R.D. 2000. Technical description of the stock synthesis assessment program. NOAA Technical Memorandum NMFS-NWFSC-43.

NMFS. 2008. Final Environmental Assessment for Amendment 24 to the Fishery Management Plan for Bering Sea/Aleutian Islands King and Tanner Crabs to Revise Overfishing Definitions.

Peacock, J.A. 1983. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society.* 202: 615-627.

Punt, A.E., Buckworth, R.C., Dichmont, C.M. and Y. Ye. 2009. Performance of methods for estimating size–transition matrices using tag–recapture data. *Marine and. Freshwater Research* 60:168-182.

Sullivan, P.J., Parma, A.M. and W.J. Clark. 1999. The Pacific halibut stock assessment of 1997. International Pacific Halibut Commission Scientific Report No. 79.

Thompson, G.G. 1992. Management advice from a simple dynamic pool model. *Fishery Bulletin* 90: 552-560.

Thompson G.G. 1993. A proposal for a threshold stock size and maximum fishing mortality rate. *Canadian Special Publications in Fisheries and Aquatic Sciences* 120: 303–320.

Turnock, B.J. and L.J. Rugolo. 2008. Appendix A to the Stock assessment of eastern Bering Sea snow crab.

Williams, E.H. and T.J. Quinn II. 1998. A parametric bootstrap of catch-age compositions using the Dirichlet distribution. pp. 371-384. In: Fishery stock assessment models. F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C. I. Zhang (eds), Alaska Sea Grant Program Report No. AK-SG-98-01, University of Alaska Fairbanks

Table 1. Five-Tier System for setting overfishing limits for crab stocks. The tiers are listed in descending order of information availability (Source: NMFS (2008)).

Information available	Tier	Stock status level	F_{OFL}
B, B_{MSY}, F_{MSY}, and pdf of F_{MSY}	1	a. $\frac{B}{B_{msy}} > 1$	$F_{OFL} = \mu_A$ =arithmetic mean of the pdf
		b. $\beta < \frac{B}{B_{msy}} \le 1$	$F_{OFL} = \mu_A \frac{B/B_{msy} - \alpha}{1 - \alpha}$
		c. $\frac{B}{B_{msy}} \le \beta$	*Directed fishery F = 0* $F_{OFL} \le F_{MSY}$†
B, B_{MSY}, F_{MSY}	2	a. $\frac{B}{B_{msy}} > 1$	$F_{OFL} = F_{msy}$
		b. $\beta < \frac{B}{B_{msy}} \le 1$	$F_{OFL} = F_{msy} \frac{B/B_{msy} - \alpha}{1 - \alpha}$
		c. $\frac{B}{B_{msy}} \le \beta$	*Directed fishery F = 0* $F_{OFL} \le F_{MSY}$†
B, $F_{35\%}{}^*$, $B_{35\%}{}^*$	3	a. $\frac{B}{B_{35\%*}} > 1$	$F_{OFL} = F_{35\%}$ *
		b. $\beta < \frac{B}{B_{35\%}*} \le 1$	$F_{OFL} = F^*{}_{35\%} \frac{\frac{B}{B^*{}_{35\%}} - \alpha}{1 - \alpha}$
		c. $\frac{B}{B_{35\%}*} \le \beta$	*Directed fishery F = 0* $F_{OFL} \le F_{35\%}$†
B, *M*, $B_{msy^{prox}}$	4	a. $\frac{B}{B_{msy^{prox}}} > 1$	$F_{OFL} = \gamma M$
		b. $\beta < \frac{B}{B_{msy^{prox}}} \le 1$	$F_{OFL} = \gamma M \frac{B/B_{msy^{prox}} - \alpha}{1 - \alpha}$
		c. $\frac{B}{B_{msy^{prox}}} \le \beta$	*Directed fishery F = 0* $F_{OFL} \le F_{\gamma M}$†
Stocks with no reliable estimates of biomass or M.	5		OFL = average catch from a time period to be determined, unless the SSC recommends an alternative value based on the best available scientific information.

*35% is the default value unless the SSC recommends a different value based on the best available scientific information.

† An $F_{OFL} \le F_{MSY}$ or proxy F_{MSY} will be determined in the development of the rebuilding plan for that stock.

Table 1. Guide for understanding the five-tier system (source: NMFS (2008)).

- F_{OFL} — the instantaneous fishing mortality (F) from the directed fishery that is used in the calculation of the overfishing limit (OFL). F_{OFL} is determined as a function of:
    - F_{MSY} — the instantaneous F that will produce MSY at the MSY-producing biomass
        - A proxy of F_{MSY} may be used; e.g., $F_{x\%}$, the instantaneous F that results in x% of the equilibrium spawning per recruit relative to the unfished value
    - B — a measure of the productive capacity of the stock, such as spawning biomass or fertilized egg production.
        - A proxy of B may be used; e.g., mature male biomass
    - B_{MSY} — the value of B at the MSY-producing level
        - A proxy of B_{MSY} may be used; e.g., mature male biomass at the MSY-producing level
    - β — a parameter with restriction that $0 \leq \beta < 1$.
    - α — a parameter with restriction that $0 \leq \alpha \leq \beta$.
- The maximum value of F_{OFL} is F_{MSY}. $F_{OFL} = F_{MSY}$ when $B > B_{MSY}$.
- F_{OFL} decreases linearly from F_{MSY} to $F_{MSY} \cdot (\beta-\alpha)/(1-\alpha)$ as B decreases from B_{MSY} to $\beta \cdot B_{MSY}$
- When $B \leq \beta \cdot B_{MSY}$, F = 0 for the directed fishery and $F_{OFL} \leq F_{MSY}$ for the non-directed fisheries, which will be determined in the development of the rebuilding plan.
- The parameter, β, determines the threshold level of B at or below which directed fishing is prohibited.
- The parameter, α, determines the value of F_{OFL} when B decreases to $\beta \cdot B_{MSY}$ and the rate at which F_{OFL} decreases with decreasing values of B when $\beta \cdot B_{MSY} < B \leq B_{MSY}$.
    - Larger values of α result in a smaller value of F_{OFL} when B decreases to $\beta \cdot B_{MSY}$.
    - Larger values of α result in F_{OFL} decreasing at a higher rate with decreasing values of B when $\beta \cdot B_{MSY} < B \leq B_{MSY}$.

Table 2. BSAI crab stocks and their Tier assignments in 2009.

Stock	**Tier**
Bristol Bay red king crab	3
Eastern Bering Sea snow crab	3
Eastern Bering Tanner crab	4
Pribilof Island red king crab	4
Pribilof Island blue king crab	4
St Matthew blue king crab	4
Norton Sound red king crab	4
Aleutian Islands golden king crab	5
Pribilof Islands golden king crab	5
Adak red king crab	5

Appendix A : Participants

Invited Speakers
Catherine Dichmont, CSIRO Marine and Atmospheric Research
Martin Dorn, Alaska Fisheries Science Center, NMFS/NOAA
Peter Hulson, University of Alaska
Jim Ianelli, Alaska Fisheries Science Center, NMFS/NOAA
Terry Quinn, University of Alaska (NFPFMC SSC)

Crab Plan Team members
Bill Bechtol, University of Alaska
Forrest Bowers. Alaska Department of Fish and Game
Wayne Donaldson, Alaska Department of Fish and Game
Bob Foy, Alaska Fisheries Science Center, NMFS/NOAA
Brian Garber-Yonts, Alaska Fisheries Science Center, NMFS/NOAA
Josh Greenburg, University of Alaska
Ginny Eckert, University of Alaska
Gretchen Harrington, NMFS, Alaska Region
Doug Pengilly, Alaska Department of Fish and Game
André Punt, University of Washington
Lou Rugolo, Alaska Fisheries Science Center, NMFS/NOAA
Herman Savikko, Alaska Department of Fish and Game
Shareef Siddeek, Alaska Department of Fish and Game
Diana Stram, North Pacific Fishery Management Council
Jack Turnock, Alaska Fisheries Science Center, NMFS/NOAA

Other participants
Anne Hollowed, Alaska Fisheries Science Center, NMFS/NOAA (NFPFMC SSC)
Clayton Jernigan, NOAA General Council
Doug Kinzey, University of Washington
Pat Livingston, Alaska Fisheries Science Center, NMFS/NOAA (NFPFMC SSC)
James Murphy, University of Washington / Alaska Fisheries Science Center, NMFS/NOAA
Nick Sagalkin, Alaska Department of Fish and Game
Jack Tagart, Tagart Consulting (reprentative of BSFRF)
Doug Woodby, Alaska Department of Fish and Game
Jie Zheng, Alaska Department of Fish and Game

Appendix B: Workshop Agenda

Wednesday May 13		
Administration	8:30 am	• Introductions • Additions to draft agenda and approval of agenda
Stock assessment reporting:		
Stock Assessment TOR	8:45 am	• Punt presentation (30 minutes) • Discussion / modifications • ACL / OFL needs
BREAK	*10:30*	
Stock Assessment TOR	10:45am	• Stock-specific actions • Data rich – snow crab • Data moderate – AI Golden king crab • Data moderate – Norton Sound red king crab
LUNCH	*12:00 pm*	
Data weighting and diagnostics:		
Practices in other assessments	1:00 pm	• Hulson overview (30 minutes) • Ianelli presentation (EBS pollock) (20 minutes) • Dorn presentation (GOA Pollock) (20 minutes) • Dichmont presentation (Australia) (20 minutes)
BREAK	*2:45pm*	
Initial Recommendations	3:00 pm	• Group discussion – what is appropriate for crab • Initial recommendations – data weighting • Initial recommendations – diagnostics • Workplan for overnight analyses
Thursday May 14		
Reprise	8:30am	• Results of overnight analyses • Snow crab (Turnock) • Red king crab (Zheng) • AI Golden king crab (Sideek) • Norton Sound rd king crab (Zheng)
BREAK	*10:30*	
Final recommendations	10:45 am	• Synthesis of examples • Final recommendations – data weighting • Final recommendations – fit diagnostics
LUNCH	*11:45 am*	
Overfishing levels for Tier 4 stocks (calculating Gamma):		
Background and history	12:45 pm	• Quinn presentation (background) (20 minutes) • Current approach (Stram / Punt?) (20 minutes) • Likely stocks for Tier 4 (group)
Proxy approaches to estimating $F_{x\%}$	1:45pm	• Maturity • Selectivity • Natural mortality • Growth
BREAK	*2:45pm*	
Reprise	3:00 pm	• Recommendations
Conclusions	4:00 pm	• Overview of recommendations (Punt) • Plans for September CPT meeting
ADJOURN	4:00 pm	

Appendix C : A Guide to the Preparation of Bering Sea and Aleutian Islands Crab SAFE Report Chapters

A chapter should be produced for the SAFE report for each crab stock, and should include all sections listed in the "Outline of SAFE Report Chapters" below. This Outline is intended to provide a consistent structure and logical flow for stock assessments; using the numbering system outlined below will help to standardize the SAFE document and make the review process for assessments more straightforward. Some variation from this outline is permissible if warranted by limitations of data, analytical methods, or other extenuating circumstances; major deviations from the suggested report structure should, however, be justified. Many of the items under Section E are not appropriate for stocks in Tier 5 (see Table 1 of this Appendix for a list of sections needed for different types of assessments). It is particularly important that all of the items listed under "Calculation of the OFL" be included to the maximum extent possible, in that many of these are critical to the fishery management process. Careful consideration should be given to all applicable SSC and CPT comments from the previous assessment(s).

Important notes:

- This guide does not provide details on what is needed regarding ABCs and ACLs and will need to be modified once these details become available.
- Dates should be specified as “2008” for the 2008 calendar year and “2008/09” for the 2008/09 fishing year. By default crab assessments are based on fishing years, but the notation 2xxx/yy should nevertheless be adopted.
- Fishing mortality values (*F*) are always full selection fishing mortalities (the *F* at fishing selection equal to 1.0).

Outline of SAFE Report Chapters

Title page and list of preparers

Executive Summary

1. Stock: species/area.
2. Catches: trends and current levels.
3. Stock biomass: trends and current levels relative to virgin or historic levels, description of uncertainty.
4. Recruitment: trends and current levels relative to virgin or historic levels.
5. Management performance: a table showing estimates of mature male biomass (at the time of mating), overfishing levels (OFL and MSST), TACs, retained catch and discards in all fisheries; show results from 2005/06 to the current year (Table 2 of this Appendix lists examples of how these tables should be constructed for stocks in each Tier)
6. Basis for the OFL: Table listing estimates of *M*, Tier level, current mature male biomass (MMB, at the time of mating), B_{MSY} (or the proxy thereof) and the basis for the calculation of B_{MSY}, current mature male biomass relative to B_{MSY} (or its proxy), γ, and the basis for calculating average catch; show from 2008/09 to the current year (Table 3 of this Appendix lists examples of how these tables should be constructed for stocks in each Tier).
7. A summary of the results of any rebuilding analyses: table showing the year by which rebuilding is expected to occur, the rebuilding time period, the catch for the next fishing year and probability of recovery to the proxy for B_{MSY} for a range of harvest

strategies (including one for which the probability of recovery within the rebuilding period is 0.5).

A. Summary of Major Changes

1. Changes (if any) to the management of the fishery.
2. Changes to the input data (e.g. specify any new data sources and which data sources have been updated)
3. Changes (if any) to the assessment methodology.
4. Changes to the assessment results, including projected biomass, TAC/GHL, total catch (including discard mortality in all fisheries and retained catch), and OFL.

B. Responses to SSC and CPT Comments

1. Responses to the most recent two sets of SSC and CPT[2] comments on assessments in general (for each comment that is addressed in the main text, list the comment and give name of the section where it is discussed; if the SSC or CPT did not make any comments on assessments in general, say so).
2. Responses to the most recent two sets of SSC and CPT[2] comments specific to the assessment (for each comment that is addressed in the main text, list the comment and give the name of section where it is discussed; if the SSC or CPT did not make any comments specific to the assessment, say so).

All comments relevant to the assessment and crab assessments in general must be listed. If a comment has not been addressed in the assessment, the comment should be listed and the reasons for not addressing it must be provided.

C. Introduction

1. Scientific name.
2. Description of general distribution (including a map, showing the stock boundary and, if possible, the actual distribution).
3. Evidence of stock structure, if any.
4. Description of life history characteristics relevant to stock assessments (e.g., special features of reproductive biology).
5. Brief summary of management history. A complete summary of the management history will be provided in the ADF&G Area Management Report appended to the annual SAFE.

D. Data (Items in this section should be presented primarily in tabular form.)

1. Summary of new information (the section should essentially repeat the information provided under Section A.2).
2. Data which should be presented as time series, separately by sex and, depending on the assessment also by maturity state and shell condition (table headers should indicate when the data were extracted, and the source for the data; years should be reported as fishing year 2xxx/yy or calendar year, depending on the fishery concerned):
    a. Total catch, partitioned by strata used in the assessment model, if any.
    b. Information on bycatch and discards. Non-retained catches and discards should ideally be reported using the categories in Table 4 to this Appendix (the table

[2] For an assessment in May, these comments will be from the SSC and CPT meetings in May and September of the previous year. For an assessment in September, these comments will be from the SSC and CPT meetings in May of the current year and September of the previous year.

header should specify the mortality rates applied to discards and bycatch, and whether the values in the table have had these mortality rates applied or not).

   c. Catch-at-length (with sample sizes) for fisheries, bycatch, and discards.
   d. Survey biomass estimates (with measures of uncertainty).
   e. Survey catch-at-length (with sample sizes), as appropriate.
   f. Other time series data (e.g., predator abundance, fishing effort).
3. Data which may be aggregated over time:
   a. Growth-per-molt; frequency of molting, etc. (by sex and perhaps maturity state)
   b. Weight-at length or weight-at-age (by sex).
4. Information on any data sources that were available, but were excluded from the assessment.

Notes:

i. Information on length-composition may be more appropriately presented in the form of plots, especially for assessments for which there in a substantial amount of such data.

ii. The reported samples sizes should reflect the actual number of samples; information on the sample sizes assumed when fitting any population models should also be reported.

E. Analytic Approach

1. History of modeling approaches for this stock

In addition to summarizing how assessments methods have changed over time, include a summary of CIE review comments from past reviews and how those comments have been taken into account.

2. Model Description

a. Description of overall modeling approach (e.g., age/size structured versus biomass dynamic, maximum likelihood versus Bayesian). If the model has not been published in its current form, its equations should be listed in full in an Appendix. It there is a technical Appendix, Items b-f below should be included in the appendix, and only a short description of the model and its estimation scheme needs to be included in this section. Specify when the fishery is assumed to occur and, if necessary, provide a table which lists the assumed time of the fishery for each year of the assessment period.

b. Reference for software used (e.g., Synthesis, AD Model Builder).

c. List and description of all likelihood components.

d. Description of how the state of the population at the start of the first year of the assessment period is determined and the size-range that the model covers.

e. Parameter estimation framework:

   i. List all of the parameters which are estimated outside of the assessment (e.g., the natural mortality rate, parameters governing the maturity schedule) along with how the values for these parameters were estimated (methods do not necessarily have to be statistical; e.g., M could be estimated by referencing a previously published value).

   ii. List all of the parameters that are estimated conditionally on those described above (e.g., full-selection fishing mortality rates, parameters governing the survey and fishery selectivity schedules, recruitments), indicate any bounds and/or priors placed on these parameters.

iii. List any constraints that imposed on the estimated parameters (including penalties on recruitment and selectivity).

f. Definition of model outputs
   i. Biomass measures (e.g., biomass of animals 50mm and larger). Indicate the assumed time of mating and that of the fishery.
   ii. Recruitment (e.g., number of males and females in the 50-55mm size-class).
   iii. Fishing mortality (e.g., full-selection F multiplied by selectivity for lengths 80 and above). Whether fishing mortality is an exploitation rate or an instantaneous rate should be reported in table headers and the text. The ideal is to report "fishing mortality" as the fully-selected instantaneous fishing mortality rate at the time of the fishery to enhance comparability amongst stock assessments.

g. Critical assumptions and consequences of assumption failures (for example, highlight assumptions regarding *M*, *q* and selectivity, to which assessments are often very sensitive).

h. Changes to any of the above since the previous assessment.

i. Outline of methods used to validate the code used to implement the model and whether the code is available.

3. Model Selection and Evaluation

a. Description of alternative model configurations[3], if any (e.g., alternative *M* values or likelihood weights; use a hierarchical approach where possible (e.g. asymptotic vs domed selectivities, constant vs time-varying selectivities)). The model configuration on which the previous assessment was based must be included in the set of model considered in order to retain comparability with previous assessments[4].

b. Show a progression of results from the previous assessment to the preferred base model by adding each new data source and each model modification in turn to enable the impacts of these changes to be assessed.

c. Evidence of search for balance between realistic (but possibly over-parameterized) and simpler (but not realistic) models.

d. Convergence status and convergence criteria for the base-case model (or proposed base-case model) such as randomization run results or other evidence of a search for the global best estimates.

e. Table (or plot) of the sample sizes assumed for the compositional data. There are several ways for specify input sample size, including:
   i. the number of animals actually measured;
   ii. a fixed constant (e.g. 500);
   iii. the application of bootstrapping approaches (e.g. Folmer and Pennington, 2000); and
   iv. as for i and iii, with a maximum imposed on the input sample size

   The first, third, and last of these approaches allows the input sample sizes (and hence the weight assigned to the compositional data) to reflect uneven sampling over time. The basis for specifying the input sample sizes should be justified and

[3] For Tier 5 assessments "model configuration" refers to the time period over which the mean catch is computed while for Tier 3 and 4 assessments it includes the time period used to define B_{MSY}/B_{REF}.

[4] This information should be included in the May and September versions of the assessment report. However, for ease of reading, information on model configurations considered but not adopted should be included in an appendix to the assessment report.

analyses conducted (see Section 4.4 below) to justify the final effective sample sizes.

f. Do parameter estimates for all models make sense, are they credible?
g. Description of criteria used to evaluate the model or to choose among alternative models, including the role (if any) of uncertainty.
h. Residual analysis (e.g. residual plots, time series plots of observed and predicted values or other approach). Note that residual analysis is expected for the base-case model below.
i. Evaluation of the model, if only one model is presented; or evaluation of alternative models and selection of final model, if more than one model is presented.

4. Results (best model(s))[5]

Results should be provided for all model runs that the assessment author considers sufficiently plausible that they could form the basis for management advice. Assessment authors should come to the May Crab Plan Team meeting with detailed results for all analyses conducted.

1. List of effective sample sizes, the weighting factors applied when fitting the indices, and the weighting factors applied to any penalties.
2. Tables of estimates (all quantities should be accompanied by confidence intervals or other statistical measures of uncertainty, unless infeasible; include estimates from previous SAFEs for retrospective comparisons):
    a. All parameters (include recruitments, selectivity parameters, any estimated growth parameters, catchability, etc.).
    b. Abundance and biomass time series, including spawning biomass and MMB.
    c. Recruitment time series (including average recruitment).
    d. Time series of catch divided by biomass..
3. Graphs of estimates (all quantities should be accompanied by confidence intervals or other statistical measures of uncertainty, unless infeasible):
    a. Fishery and survey selectivities, molting probabilities, and other schedules depending on parameter estimates.
    b. Estimated male, female, mature male, total and effective mature biomass time series (indicate the proxy for B_{MSY} on the relevant plots).
    c. Estimated full selection F over time.
    d. Estimated fishing mortality versus estimated spawning stock biomass, including applicable OFL and maximum F_{target} definitions for the stock (see, for example, Fig. 54 of Turnock and Rugolo, 2008). Graphs of this type are useful to evaluate management performance.
    e. Fit of a stock-recruitment relationship, if feasible.
4. Evaluation of the fit to the data:
    a. Graphs of the fits to observed and model-predicted catches (retained catch and discards), including model-predicted of catches and discards for all years to allow discards to be inferred for years for which data are not available.
    b. Graphs of model fits to survey numbers (include confidence intervals for the data and model predictions).
    c. Graphs of model fits to catch proportions by length (e.g. using bubble and/or line plots).

[5] There may be several "best" models in the May assessment draft, but there should be one "best" model in the September assessment draft.

   d. Graphs of model fits to survey proportions by length (e.g. using bubble and/or line plots).
   e. Marginal distributions for the fits to the compositional data.
   f. Plots of implied versus input effective sample sizes and time-series of implied effective sample sizes.
   g. Tables of the RMSEs for the indices (and a comparison with the assumed values for the coefficients of variation assumed for the indices).
   h. Quantile-quantile (q-q) plots and histograms of residuals (to the indices and compositional data) to justify the choices of sampling distributions for the data.
5. Retrospective and historic analyses (retrospective analyses involve taking the "best" model and truncating the time-series of data on which the assessment is based; a historic analysis involves plotting the results from previous assessments).
   a. Retrospective analysis (retrospective bias in base model or models).
   b. Historic analysis (plot of actual estimates from current and previous assessments).
6. Uncertainty and sensitivity analyses (this section should highlight unresolved problems and major uncertainties, along with any special issues that complicate scientific assessment, including questions about the best model, etc.):
   a. The best approach for describing uncertainty depends on the situation. Possible approaches (not mutually exclusive) are:
      i. Sensitivity analyses (tables or figures) that show ending biomass levels, OFLs, and/or likelihood component values obtained while systematically varying (e.g. halving and doubling) the emphasis factors for each type of data (and penalty) in the model.
      ii. Likelihood profiles for parameters or biomass levels.
      iii. CVs for biomass or OFL estimated by bootstrap, the delta method or Bayesian methods.
      iv. Subjective appraisal of the magnitude and sources of uncertainty.
      v. Retrospective and historic analyses (see above).
      vi. Comparison of alternate models and or assumptions (i.e. model structure uncertainty, as evaluated in Section E.3 of this Appendix).
   b. It is important that some qualitative or quantitative information about relative probability be stated if a range of model runs (e.g., based on CV's or alternative assumptions about model structure or recruitment) is used to depict uncertainty. It is important to state that all scenarios (or all scenarios between the bounds depicted by the runs) are equally likely if no statements about relative probability can be made.
   c. Simulation results.

F. Calculation of the OFL

1. Specification of the Tier level and stock status level for computing the OFL, along with the basis for the selection. For Tier 4 and 5 stocks, the rationale for the time period used to define B_{REF} (Tier 4) and the average retained catch used to compute the OFL (Tier 5) needs to be specified.
2. List of parameter and stock size estimates (or best available proxies thereof) required by limit and target control rules specified in the fishery management plan.
3. Specification of the OFL:
   a. Provide the equations (from Amendment 24) on which the OFL is to be based, including the equations used to project discard and bycatch by sex (the

mathematical specifications for this need to be documented in a peer-reviewed publication or in a technical appendix).

b. Basis for projecting MMB to the time of mating (the mathematical specifications for this need to be documented in a peer-reviewed publication or in a technical appendix).

c. Specification of F_{OFL}, OFL, and other applicable measures (if any) relevant to determining whether the stock is overfished or if overfishing is occurring (such as B_{REF}, $B_{35\%}$). Include estimates from the present assessment and the assessments since 2006/07. Table 2 of this Appendix lists examples of tables for Tiers 3, 4 and 5.

4. Recommendation for F_{OFL}, OFL total catch (or OFL retained catch) for the coming year. List the OFLs by sector (retained catch, discard in the directed fishery, bycatch in other crab fisheries, the groundfish fishery, etc.), where appropriate.

G. Rebuilding Analyses

Rebuilding analyses should be provided for stocks which are currently under a rebuilding plan.

1. Definition of recovery (including the definition of the proxy for B_{MSY}, the number of years that the biomass needs to exceed the proxy for B_{MSY} for the stock to be recovered).
2. Year in which the rebuilding plan started and the year by which the stock should be recovered to the proxy for B_{MSY}.
3. Specification of the approach used to project the model forward (e.g. assumptions about parameter uncertainty; future recruitment and selectivity; and how discards and bycatch are computed given fishing mortality on mature males).
4. Projections under different levels of fishing mortality on mature males to evaluate the probability of recovery to the proxy for B_{MSY} over time. Results should be produced for (a) no targeted fishing, (b) probabilities of recovery of 0.5, 0.6, 0.7 and 0.8, and (c) a harvest strategy corresponding to 75% of the F_{OFL}.
5. Tables of total catch, retained catch, and probability of recovery against time for the rebuilding strategies listed under 4).
6. A graph of the annual status of the stock relative to the B_{MSY} and MSST from the start of the rebuilding period to the present.

H. Data Gaps and Research Priorities

Information which could feasibly be collected and analyses which should be undertaken to improve the assessment should be included in this section. Ideally, data collection and analysis needs should be listed in priority order.

I. Ecosystem Considerations

Discussion of any ecosystem considerations (e.g., relationships with species listed under the ESA, prohibited species concerns, bycatch issues, refuge areas, and gear considerations).

The following subsections should provide information on how various ecosystem factors might be influencing the stock or how the fishery might be affecting the ecosystem and what data gaps might exist that prevent assessing such effects.

1. Ecosystem Effects on Stock

There are several factors that should be considered for each stock in this subsection. These include:

1. Prey availability/abundance trends (historically and in the present and foreseeable future). These prey trends could affect growth or survival of a target stock.
2. Predator population trends (historically and in the present and foreseeable future). These trends could affect mortality rates over time.
3. Changes in habitat quality (historically and in the present and foreseeable future). These would primarily be changes in the physical environment such as temperature, currents, or ice distribution that could affect stock migration and distribution patterns, recruitment success, or direct effects of temperature on growth.

2. Fishery Effects on the Ecosystem

In this section the following factors should be considered:

1. Fishery-specific bycatch of HAPC biota (in particular, species common to *YourFishery*), marine mammals and birds, and other sensitive non-target species.
2. Fishery-specific concentration of target catch in space and time relative to predator needs in space and time (if known) and relative to spawning components.
3. Fishery-specific effects on amount of large size target crab.
4. Fishery-specific contribution to discards.
5. Fishery-specific effects on age-at-maturity and fecundity of the target species.
6. Fishery-specific effects on EFH non-living substrate (using gear specific fishing effort as a proxy for amount of possible substrate disturbance).

Authors should consider summarizing the results of these analyses into a table as shown below (for example):

Analysis of ecosystem considerations for *YourStock* and the *YourFishery*. The observation column should summarize the past, present, and foreseeable future trends. The interpretation column should provide details on how the trend affects the stock (ecosystem effects on the stock) or how the fishery trend affects the ecosystem (fishery effects on the ecosystem). The evaluation column should indicate whether the trend is of: *no concern, probably no concern, possible concern, definite concern,* or *unknown.*

Ecosystem effects on *YourStock*			
Indicator	Observation	Interpretation	Evaluation
Prey availability or abundance trends			
Zooplankton	Stomach contents, ichthyoplankton surveys, changes mean wt-at-age	Stable, data limited	Unknown
Predator population trends			
Marine mammals	Fur seals declining, Steller sea lions increasing slightly	Possibly lower mortality on pollock	No concern
Birds	Stable, some increasing some decreasing	Affects young-of-year mortality	Probably no concern
Fish (Skate, flatfish, pollock, Pacific cod, halibut)	Stable to increasing	Possible increases to pre-recruit crab mortality	Probably concern (young of the year is not dealt within the model?)
Changes in habitat quality			
Temperature regime	Cold years pollock and other demersal fish distribution towards NW on average	Likely to affect surveyed stock	No concern (dealt with in model)
Winter-spring environmental conditions	Affects pre-recruit crab survival	Probably a number of factors	Causes natural variability
Production	Fairly stable nutrient flow from upwelled BS Basin	Inter-annual variability low	No concern

***YourFishery* effects on ecosystem**			
Indicator	Observation	Interpretation	Evaluation
Fishery contribution to bycatch			
Prohibited species	Stable, heavily monitored	Minor contribution to mortality	No concern
Forage (including herring, Atka mackerel, cod, and pollock)	Stable, heavily monitored	Bycatch levels small relative to forage biomass	No concern
HAPC biota	Low bycatch levels of (spp)	Bycatch levels small relative to HAPC biota	No concern
Marine mammals and birds	Very minor direct-take	Safe	No concern
Sensitive non-target species	Likely minor impact	Data limited, likely to be safe	No concern
Fishery concentration in space and time	Generally patchy	Mixed potential impact (fur seals vs Steller sea lions)	Possible concern
Fishery effects on amount of large size target fish	Depends on highly variable year-class strength	Natural fluctuation	Probably no concern
Fishery contribution to discards and offal production	Decreasing	Improving, but data limited	Possible concern
Fishery effects on age-at-maturity and fecundity		NA	Possible concern

J. Literature Cited

Include citations that are relevant to understanding the stock and its status, but are not cited in the report in a special “extra references” section.

Table 1. Requirements for assessments by Tier.

Report Section	**Tiers 1-3; Tier 4 (with assessment)**	**Tier 4 (no assessment)**	**Tier 5**
Executive Summary	Yes	Yes	Yes
A. Summary of Major Changes	Yes	Yes	Yes
B. Responses to SSC and CPT comments	Yes	Yes	Yes
C. Introduction	Yes	Yes	Yes
D. Data	Yes	Yes[1]	Yes[2]
E. Analytical Approach	Yes	Yes[3]	Yes[3]
F. Calculation of the OFL	Yes	Yes	Yes
G. Rebuilding Analyses	Yes[4]	Yes[4]	Yes[4]
H. Data Gaps and Research Priorities	Yes	Yes	Yes
I. Ecosystem Considerations	Yes	Yes	Yes
J. Literature Cited	Yes	Yes	Yes

1 – Items 2c, 2e need not be reported in full
2 – Items 2c -2e need not be reported in full
3 – Limited to plots of survey data and catches
4 – Only for stocks under rebuilding

Table 2. Examples of summary tables of management performance by Tier level (the table is structured for an assessment conducted in September 2009)

(a) Stocks in Tiers 1-3 and those in Tier 4 for which there is an agreed assessment model

Year	MSST	Biomass (MMB)	TAC	Retained Catch	Total Catch	OFL
2005/06		100[A]	60	40	58	
2006/07		120[B]	60	51	55	
2007/08	230[C]	130[C]	60	55	56	
2008/09	221[D]	219[D]	60	47	55	91
2009/10		280[D]				78

The stock was above MSST in 2008/09 and is hence not overfished. Overfishing did not occur during the 2008/09 fishing year.

Notes:
A – Calculated from the assessment reviewed by the Crab Plan Team in September 2006
B – Calculated from the assessment reviewed by the Crab Plan Team in September 2007
C – Calculated from the assessment reviewed by the Crab Plan Team in September 2008
D – Calculated from the assessment reviewed by the Crab Plan Team in September 2009

(b) Stocks in Tier 4 for which there is not an agreed assessment model

Year	MSST	Biomass (MMB)	TAC	Retained Catch	Total Catch	OFL
2005/06		100[A]	60	40	58	
2006/07		120[B]	60	51	55	
2007/08	230[C]	130[C]	60	55	56	
2008/09	221[D]	219[D]	60	47	55	91
2009/10		280[D]				78

The stock was above MSST in 2008/09 and is hence not overfished. Overfishing did not occur during the 2008/09 fishing year.

Notes:
A – Based on survey data available to the Crab Plan Team in September 2006 (even though it may have been updated)
B – Based on survey data available to the Crab Plan Team in September 2007 (even though it may have been updated)
C – Based on survey data available to the Crab Plan Team in September 2008 (even though it may have been updated)
D – Based on survey data available to the Crab Plan Team in September 2009

(c) Stocks in Tier 5

Year	MSST	Biomass (MMB)	TAC	Retained Catch	Total Catch	OFL
2005/06		N/A	60	40	58	
2006/07		N/A	60	51	55	
2007/08	N/A	N/A	60	55	56	
2008/09	N/A	N/A	60	47	55	91
2009/10		N/A				78

No overfished determination is possible for this stock given the lack of biomass information. Overfishing did not occur during the 2008/09 fishing year.

Table 3. Examples of tables that summarize how the OFL was calculated (the table is structured for an assessment conducted in September 2009). The rows for 2008/09 were agreed by the Crab Plan Team in September 2008 and those for 2009/10 were agreed by the Crab Plan Team in September 2010.

(a) Stocks in Tiers 1-3 and those in Tier 4 for which there is an agreed assessment model

Year	Tier	B_{MSY}	Current MMB	B/B_{MSY} (MMB)	F_{OFL}	Years to define B_{MSY}	Natural Mortality
2008/09	3b	231	219.5	0.95	$0.15yr^{-1}$	1978/79-2008/09	$0.25yr^{-1}$
2009/10	3a	234	245.7	1.05	$0.19yr^{-1}$	1978/79-2009/10	$0.25yr^{-1}$

(b) Stocks in Tier 4 for which there is not an agreed assessment model

Year	Tier	B_{MSY}	Current MMB	B/B_{MSY} (MMB)	γ	Years to define B_{MSY}	Natural Mortality
2008/09	4b	231	219.5	0.95	1.0	1978/79-2008/09	$0.25yr^{-1}$
2009/10	4a	234	245.7	1.05	0.6	1978/79-2009/10	$0.25yr^{-1}$

(c) Stocks in Tier 5

Year	Tier	Years to define Average catch (OFL)	Natural Mortality
2008/09	5	1978/79-2008/09	$0.25yr^{-1}$
2009/10	5	1978/79-2009/10	$0.25yr^{-1}$

Table 4. Categories for which information on catches and discards should ideally be provided.

Directed pot fishery (males)
Directed pot fishery (females)
Bycatch in other crab fisheries (by sex)
Bycatch in groundfish pot (by sex)
Bycatch in groundfish trawl (by sex)
Bycatch in the scallop fishery

Appendix D : Evaluating for the Method For Evaluating Random Residuals

Figure D.1 summarizes the application of the method for evaluating whether a residual pattern is random. Results are shown for six cases. The first two cases (a and b) are cases in which the residuals are obviously mis-specified. Cases c and d show results for random residuals and cases e and f respectively show results when there are cohort effects (Fig. D1e) and year-effects (Fig. D1f). Application of the two-sample Kologorov-Smirnov test to the six cases leads to p-values of $<10^{-10}$, $<10^{-10}$, 0.8225, 0.9242, 0.06818, and 0.1229 respectively. The apparent low power for case f suggests that there may be value in developing tests for specific alternative hypotheses.

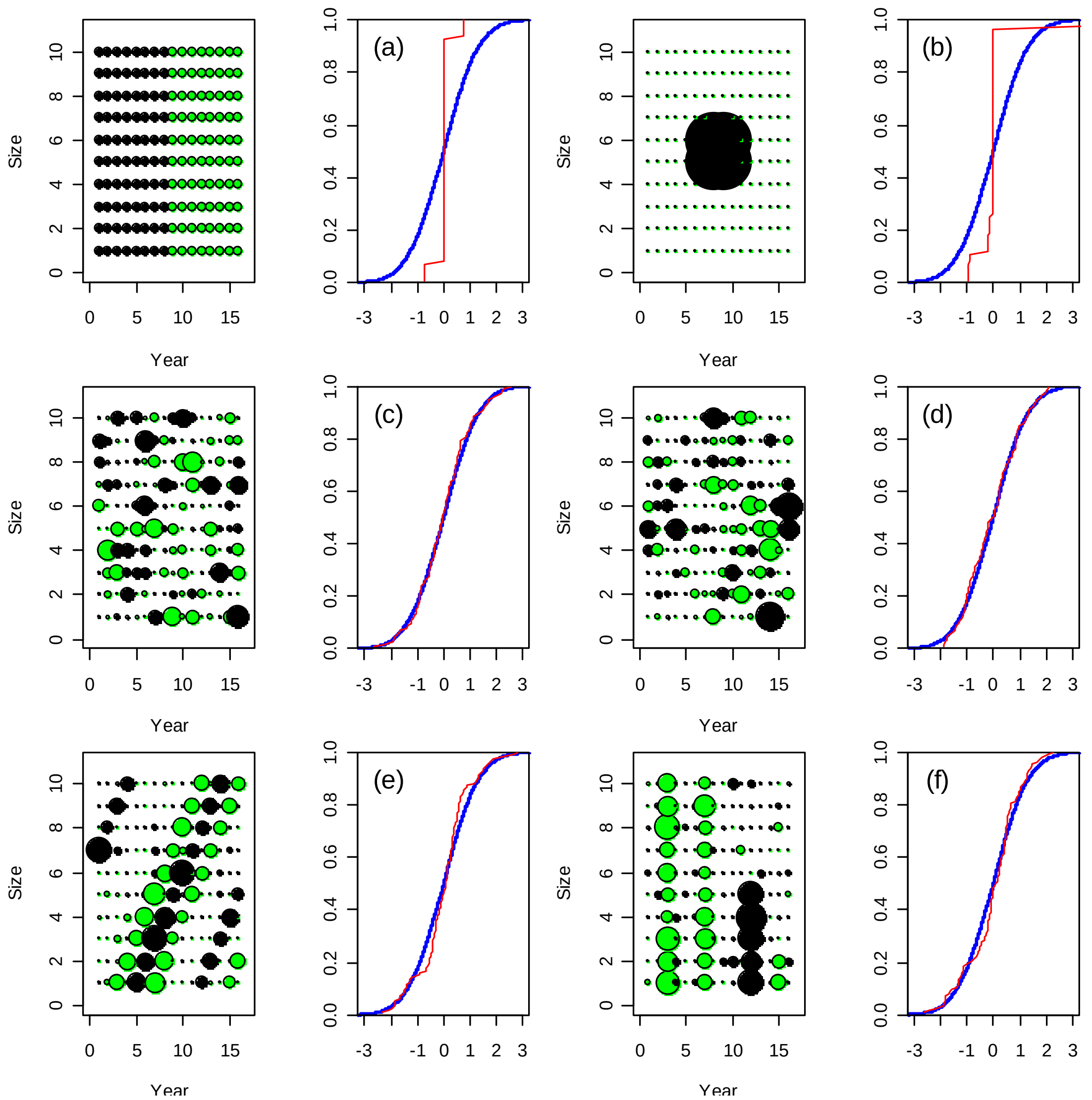

Figure D.1. Results of applying a method for evaluating whether residual patterns are random. The thick line is the null distribution and the thin line is the cumulative distribution of the test statistic.

Appendix E : Summary of Diagnostic Plots

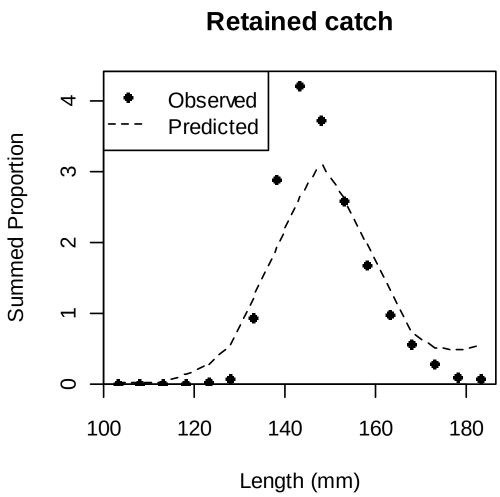

Figure E.1(a). Marginal observed and predicted catch size-compositions for AIGKC.

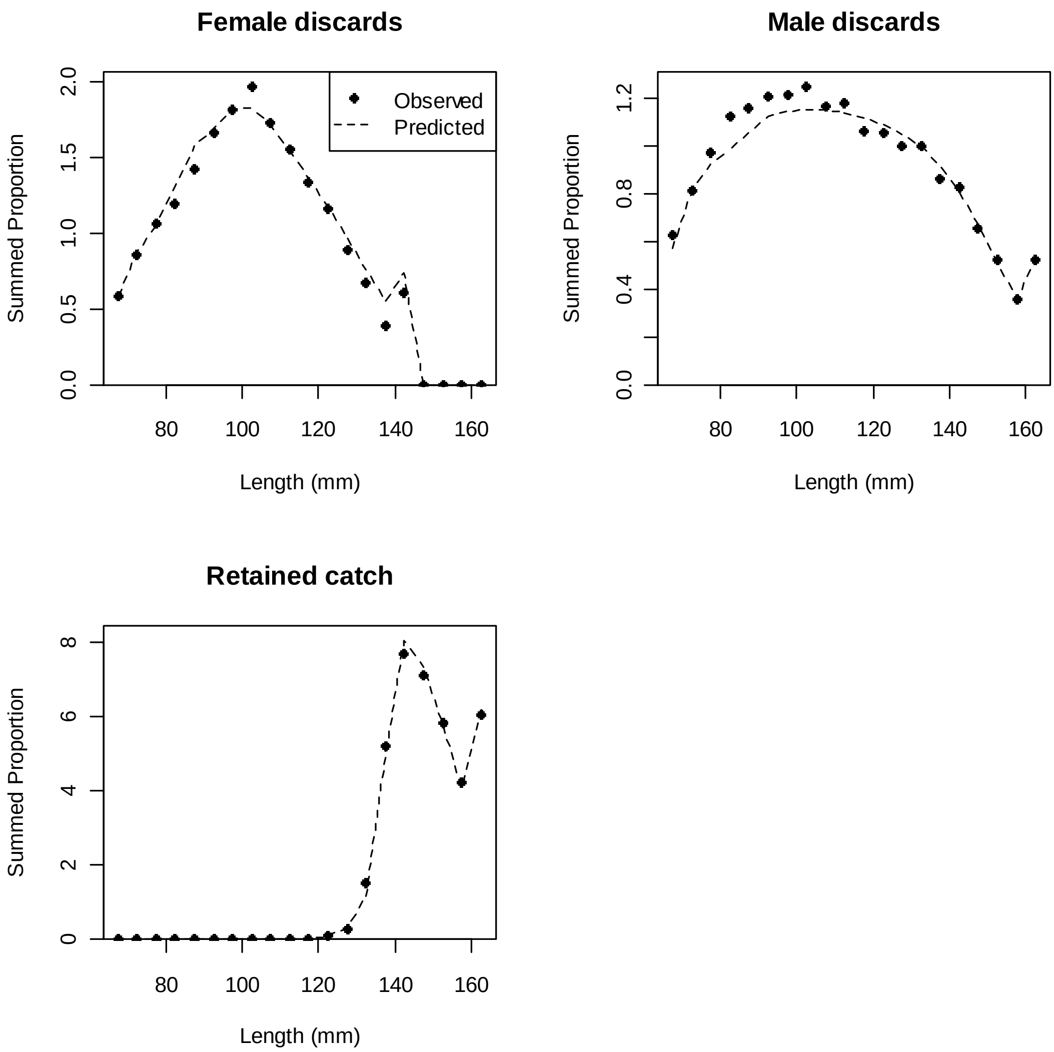

Figure E.1(b). Marginal observed and predicted catch size-compositions for BBRKC.

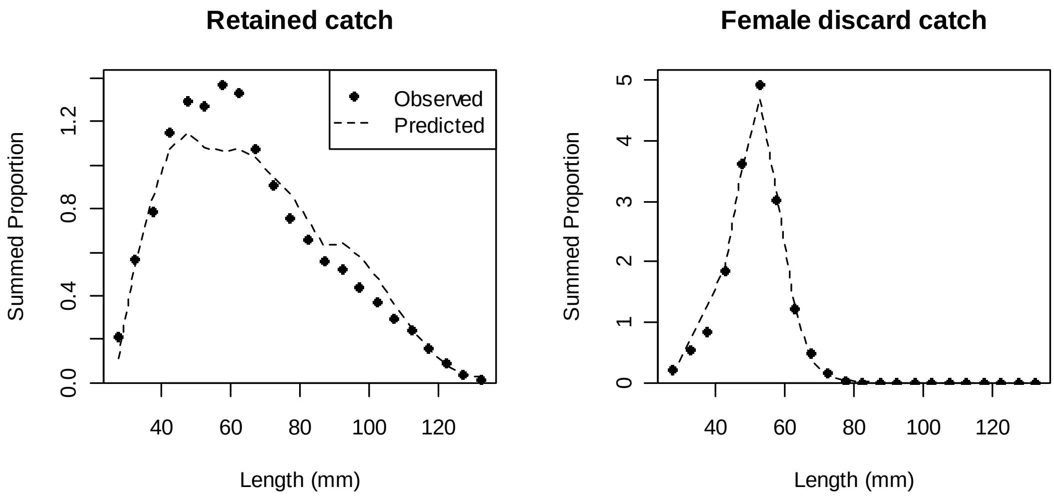

Figure E.1(c). Marginal observed and predicted catch size-compositions for snow crab.

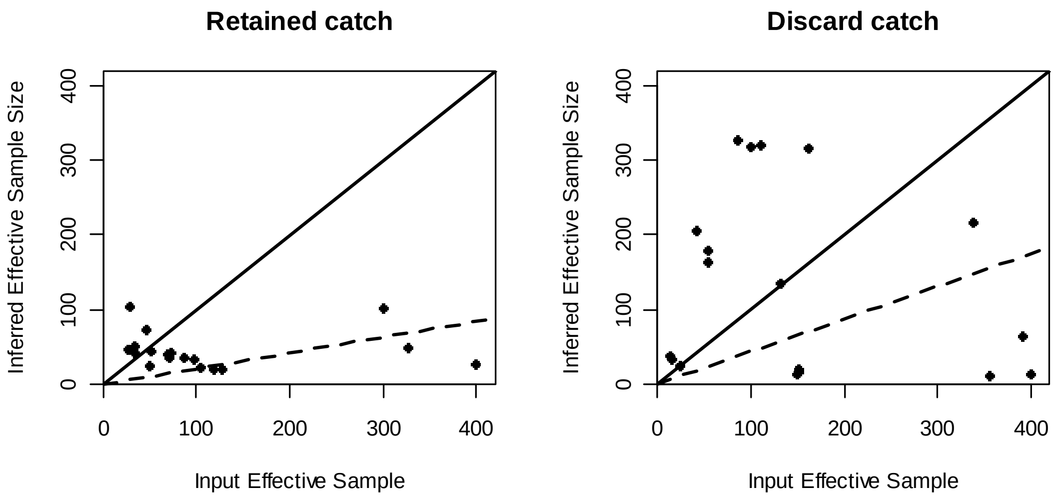

Figure E.2(a) Observed and "implied" effective sample sizes for AIGKC. The dashed line is a fit to the data and the solid line is the 1-1 line.

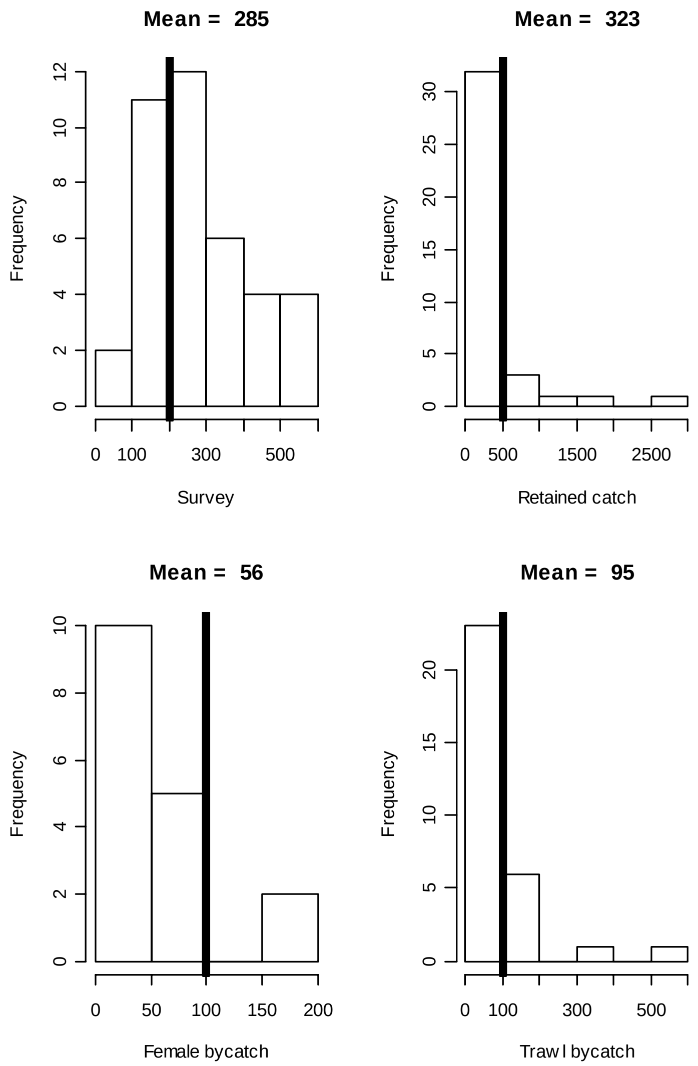

Figure E.2(b) “Implied” sample sizes (bars) and the input value (solid vertical line) for BBRKC.

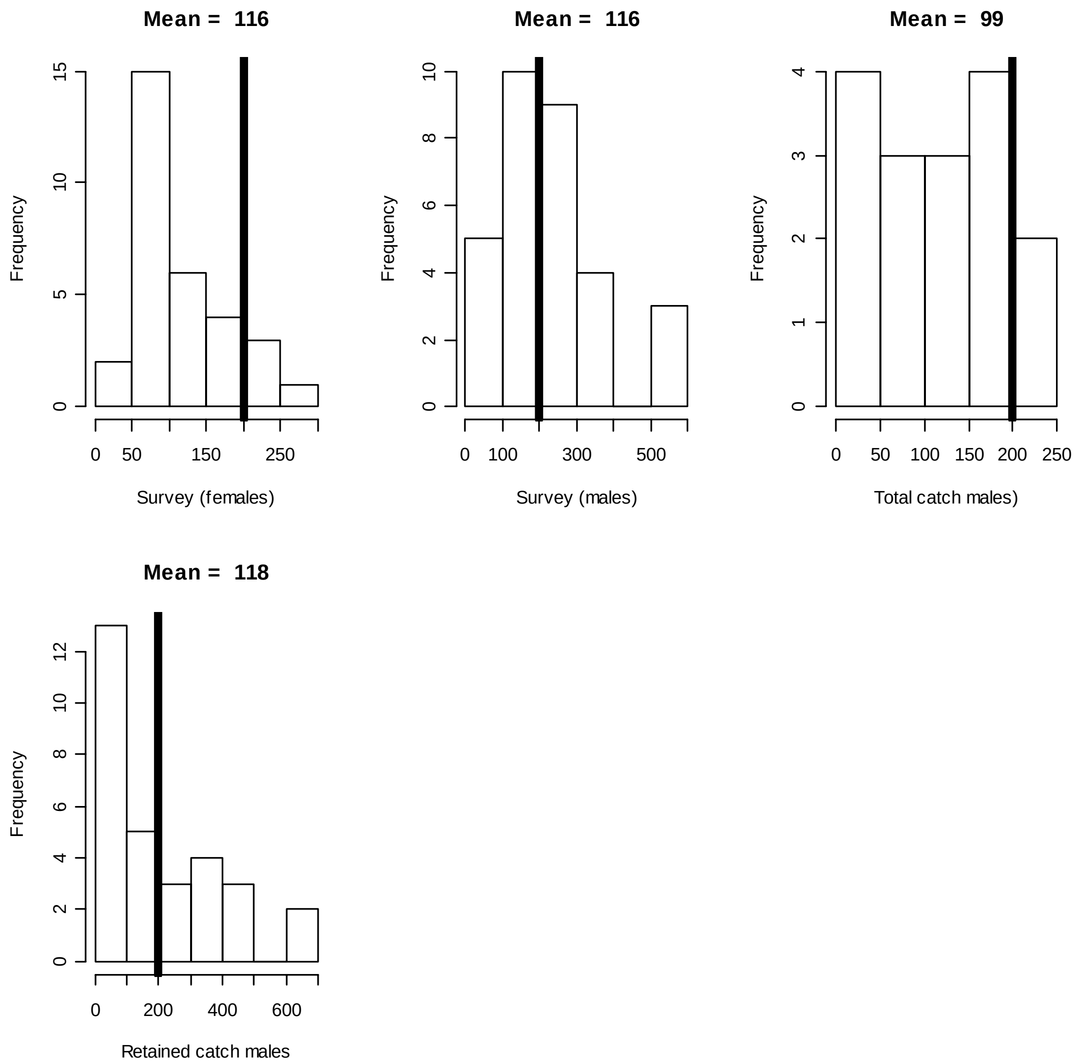

Figure E.2(c) “Implied” effective sample sizes (bars) and the input value (solid vertical line) for snow crab.

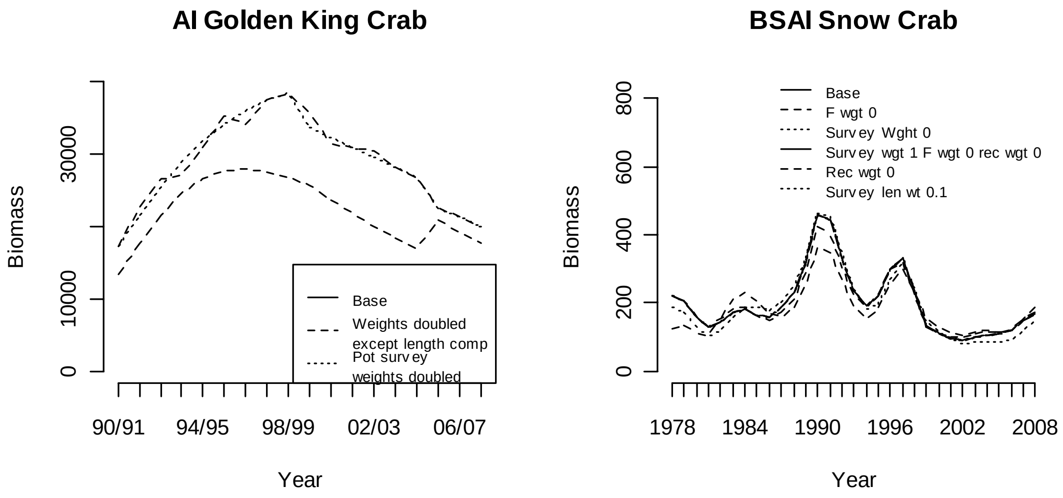

Figure E.3. Sensitivity of the time-trajectories of mature male biomass for AIGKC and snow crab to changing the weights assigned to the data sources.

Appendix F : A Surplus Production Approach to Estimating F_{MSY}

Basic approach

It is possible to estimate surplus production empirically for stocks/species for which estimates of abundance (in units of mature male biomass) are available from surveys or from stock assessments (Hilborn, 2001) using the formula:

$$S_y = B_{y+1} - B_y + C_y \tag{F.1}$$

where S_y is the surplus production (in mature male biomass) during year *y*,
B_y is the biomass (in mature male biomass) at the start of year *y*, and
C_y is the catch during year *y*.

The annual surplus production rate is defined as the ratio of the annual surplus production to the average biomass over the year, i.e. $\tilde{S}_y = 2S_y / (B_y + B_{y+1})$. If $\tilde{y}$ is the set of the years which correspond (approximately) to when the stock was close to B_{MSY} then the average value of $\tilde{S}_y$ over the years in $\tilde{y}$ is an estimate of the exploitation rate at which *MSY* is achieved.

Simulation testing

Preliminary testing of the basic approach was conducted using simulation. Simulated biomass series were generated using the equations:

$$B_{y+1} = (B_y + rB_y(1 - B_y / K) - C_y)e^{\varepsilon_y - \sigma_p^2/2} \qquad \varepsilon_y \sim N(0;\sigma_p^2) \tag{F.2a}$$

$$C_y = qE_yB_y \qquad E_y \sim e^{\eta_y} \qquad \eta_y \sim N(0;\sigma_E^2) \tag{F2b}$$

where r is the intrinsic rate of growth,
K is the carrying capacity,
q is catchability,
σ_p is the extent of process error, and
σ_E is the extent of variability in effort.

The catches were assumed to be measured exactly while the estimates of biomass were assumed to be subject to log-normal measurement error, i.e. $\hat{B}_y = Be^{\phi_y - \sigma_v^2/2}$; $\phi_y \sim N(0;\sigma_v^2)$ where σ_v determines the extent of measurement error. For the purposes of this preliminary investigation, $\tilde{y}$ was defined as the set of years for which $0.3 < \hat{B}_y / K < 0.7$. Table F.1 lists the baseline values for the parameters of the simulation model. Figure F.1 shows plots of the (true) simulated biomass relative to carrying capacity and the left panel of Figure F.2a shows the simulated distribution of estimate of F_{MSY} for the baseline values of the parameters, while Figures F.2b and F.2c shows the impact of there being no process (F.2.b) or observation (F.2c) error.

Table F.1. The baseline values for the parameters of the simulation model

Intrinsic growth rate, r	0.2
Carrying capacity, K	1000
Catchability, q	0.05
Extent of process error, σ_p	0.2
Extent of variability in effort, σ_E	0.2
Extent of survey error, σ_v	0.5

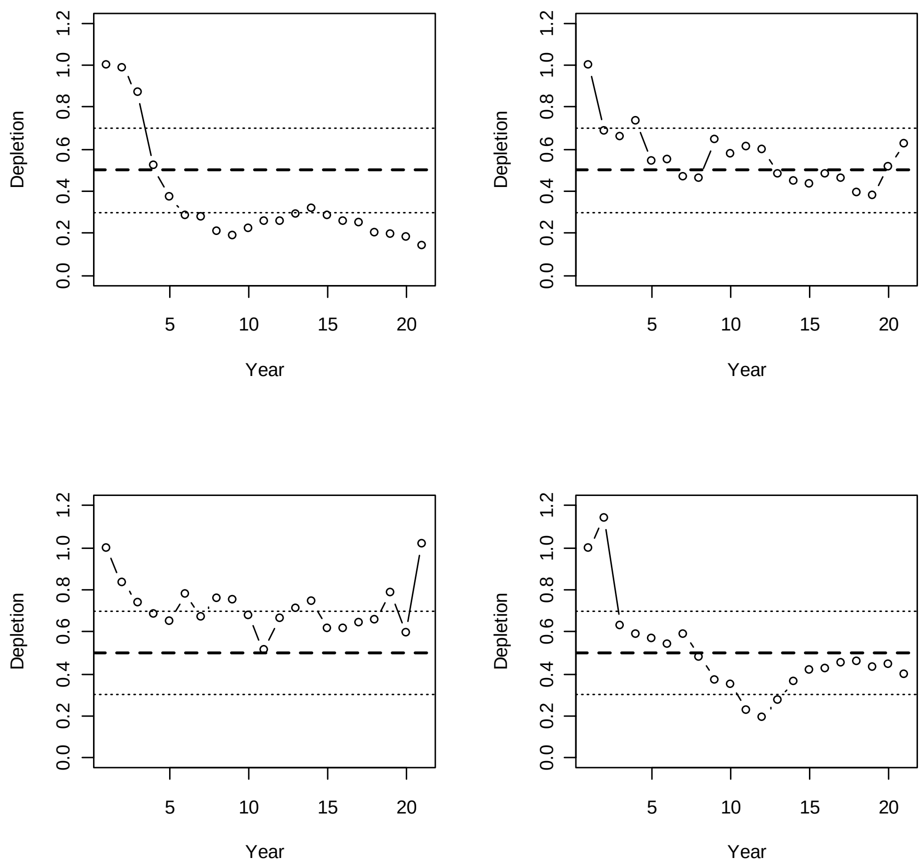

Figure F.1. True simulated biomass relative to carrying capacity (dots) and the range of values used in the calculation of the F_{MSY} (values between the dotted lines).

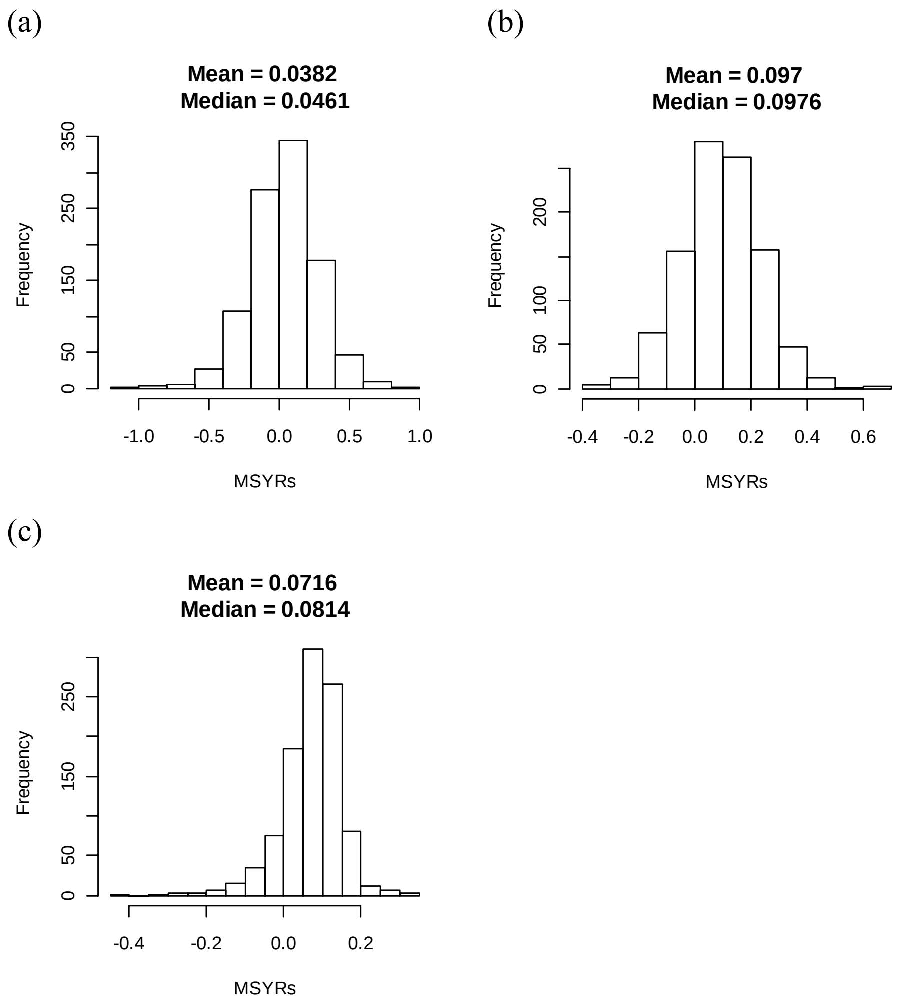

Figure F.2. Estimates of F_{MSY} from simulated data sets: (a) baseline parameter values, (b) no process error, and (c) no measurement error.

Appendix G : An Approach to Estimating Gamma

(After Turnock and Rugolo, 2008)

The F_{MSY} proxy for the control rule is F_{MSY} proxy = γM, where $\gamma = F_{35\%}/M$ so the F_{MSY} proxy = $(F_{35\%}/M) * M$. The use of $F_{35\%}$ as the F_{MSY} proxy in the control rule is equivalent to using γ, where γ is estimated as $F_{35\%}/M$. This value of $F_{35\%}$ is used with the estimated fishery selectivities to estimate the OFL. This value of γ is specific to the $F_{35\%}$ and the estimated fishery selectivities and cannot be used without those fishery selectivities, for example in the product of γ, M, and mature male biomass to estimate the total catch OFL.

Discard and retained selectivities, S_l, can be estimated using the length frequency of the observed catch, the ratio of discarded to retained numbers of crab, and the predicted catch length frequency and numbers (discard and retained) using the recent survey abundance by length projected forward to the time of the fishery. The proportion of males which are mature at length, P_l, and the vector of weights-at-length, w_l, for males can also be inferred from survey data. Given a size transition matrix **X** and a value for natural mortality, M, the population can be projected forward using the equation[6]:

$$\underline{N}_{t+1} = \mathbf{XS}\underline{N}_t + \underline{R} \tag{G.1}$$

where $\underline{N}_t$ is the numbers at length at the start of year t,

$\underline{R}$ is the recruitment by length-class (set to unity multiplied by the proportion of recruitment that occurs to each size-class),

S is a matrix with $(1 - S_l F)e^{-M}$ on the diagonal and zero elsewhere, and

F is fishing mortality

Denoting the equilibrium point of Equation G.1 as $\underline{N}^*(F)$, the mature biomass-per-recruit at the time at mating can be computed as:

$$MMB(F) = \sum_l w_l N_l^*(F)(1 - S_l F)e^{-\phi M} \tag{G.2}$$

The value for $F_{35\%}$ is selected so that $MMB(F_{35\%}) / MMB(0) = 0.35$.

[6] The derivation is based on one particular crab life history. The specific forms of Equations G.1 and G.2 will depend on, for example, the number and timing of the various fisheries.