SCIENTIFIC COMMITTEE
SEVENTH REGULAR SESSION
7-15 August 2011
Busan, Republic of Korea

INDEPENDENT REVIEW OF 2011 WCPO BIGEYE TUNA ASSESSMENT

WCPFC-SC8-2012/SA-WP-01

James Ianelli[1], Mark Maunder[2] and André E. Punt[3]

[1] Affiliate professor at the University of Washington and a stock assessment scientist with the Resource Ecology and Fisheries Management division of the Alaska Fisheries Science Center
[2] Head of the Stock Assessment Program at the Inter-American Tropical Tuna Commission
[3] Professor of Aquatic and Fishery Sciences at the University of Washington

Independent Review of 2011 WCPO Bigeye Tuna Assessment

James Ianelli, Mark Maunder, and André E. Punt

Executive Summary

1. The stock assessment for bigeye tuna in the WCPO is based on state-of-the art methods and is analytically very thorough. The analysis of raw data, where available, is more comprehensive than is common for most assessment applications.
2. The amount and quality of the data used in the assessment vary spatially and temporally, and several of the data sources appear to be in conflict, such that removing (or reweighting) some data sources can lead to qualitatively different outcomes from the assessment.
3. Previous analyses using a Pacific-wide assessment appeared to justify the current assessment approach of conducting an assessment for the WCPO only. However, recent tagging data show considerable movement of bigeye tuna between the WCPFC area and the eastern Pacific Ocean (EPO). This suggests that the assumption that migration between the WCPO and EPO need not be explicitly accounted for in the assessment needs to be re-evaluated. Given the new information, a new Pacific-wide assessment should be conducted to re-evaluate this.
4. The results of the reference model are particularly sensitive to (a) inclusion of tagging data from eastern Australia which came from a limited area (relative to the distribution of bigeye tuna in region 5), and (b) the early CPUE data for the Japanese longline fisheries. Addressing these two major sources of uncertainty should be the focus of the next assessment.
5. Given model configuration and assumptions, there is no definitive basis to select between estimating B_{MSY} based on the entire sequence of recruitment and spawning biomass estimates or more recent values. Basing catch limits on a constant fishing mortality strategy is more likely to be robust to this uncertainty than strategies which are based on the estimate of B_{MSY}. However, projected times to rebuild to the estimate of B_{MSY} should be monitored.
6. The Panel identified several areas where collection of additional data will be beneficial as well as suggestions for methodological improvements and further data analyses.

Introduction and General Issues

The Panel (see Appendix A for panel biographies) conducted a review of the 2011 assessment of bigeye tuna (BET) for the western and central Pacific, including the data inputs, the settings for the reference model and the settings for the sensitivity tests, based on the final Terms of Reference (Appendix B). Prior to the meeting Dr Shelton Harley provided an annotated version of key questions related to each ToR (included as second part of Appendix B). The Panel was provided with a set of background documents (Appendix C) prior to the meeting of the Panel, as well as MULTIFAN-CL (MFCL) input and output files, and code to view these files.

The review meeting took place between 29 April and 2 May 2012 at SPC, Noumea, New Caledonia, and was chaired by Dr André Punt. The analysts (see Appendix D) gave a presentation of the 2011 assessment and responded to questions from the Panel on the first day of the review. The Panel identified a number of requests for additional model runs and data analyses which the analysts addressed between meeting sessions (including some requests made prior to the meeting). During the subsequent days, the Panel evaluated the responses to its requests (Appendix E), and reviewed the background documents. A draft report was presented to the analysts on 2 May 2012, and finalized after the review meeting.

The Panel considered three key questions which have a fundamental impact on inferences regarding population status, as a focus for its review: (a) what aspects of the assessment determine the absolute scale of estimated population size, (b) what aspects of the assessment determine the estimated trend in population size (historically and in recent years), and (c) what aspects of the assessment determine the estimated trend in recruitment. The analyses presented to the Panel confirmed that some of the data sources were inconsistent (i.e. changing the weights assigned the various data sources led to qualitatively different outcomes).

The Panel notes that because the regions are linked through common selectivity patterns as well as assumptions regarding the relative catchabilities between the Japanese longline fleets in different regions, assumptions for one region will impact estimated abundance for other regions. Specifically, the Panel was advised by the analysts that the tagging data for region 5 imposed a bound on the biomass in this region. Given the constraints on catchability for the Japanese longline fleets, this bound impacts estimated biomass in other regions. The Panel **recommends** that the way the fisheries are linked be more fully documented in the assessment report, and the implications of such linkage be more fully evaluated.

Although the evidence warrants omitting the Australian tagging data in region 5 and the length-frequency data for the Japanese longline fisheries from the reference assessment, the Panel cautions against dropping data sources just because the model is unable to replicate them. This is because most models, even ones of this complexity, will be unable to mimic all of the data since some inconsistencies are to be expected. The rationale for the omitting these two data sources from the assessment is due to the uncertainty in their derivation (as is the case for the length frequency data for the Japanese longline fisheries) or that assumptions about population dynamics are inconsistent with the way the data were collected, as is the case for the Australian tagging data since the assumption that tags are well-mixed in region 5 after allowing for time to allow tags to mix appears incorrect. For example, should detailed documentation and evaluations of Japanese length frequency become available in the future, these data might be useful to include in the assessment provided they were consistent with other validated data on size composition (i.e., the weight frequency data for those fisheries).

The Panel requested a number of alternative model runs during the review meeting. These highlighted that the assessment outcomes were quite variable if some of the tagging data were omitted from the assessment. Sensitivity tests in future assessments should consider the impact of factors that constrain the size of and trends in biomass estimates. In the case of this assessment, the Panel identified the tagging data, in conjunction with the upper bound on the reporting rate (particularly for eastern Australia), to substantially reduce the sensitivity of model outcomes to changes to model assumptions and data.

The Panel wish to thank the SPC for hosting the meeting, the thorough background information provided prior to the review meeting, and the participants for the excellent and constructive atmosphere during the review meeting. The Panel particularly wish to thank the analysts (Nick Davies and Simon Hoyle) for their skill in addressing the many requests from the Panel quickly and for their considerable patience. The availability of results and analyses during the review meeting substantially enhanced the Panel's ability to address its ToR. In conclusion, the Panel recognizes that the current assessment, while it can be improved, is state of the art. The Panel was impressed by the comprehensive analysis of raw data which allowed it to explore a variety of model specifications and test several assumptions. The thoroughness reflected in the underlying data evaluations/preparations extended into the model sensitivities and the structural analysis.

Panel Deliberations Relative to Each TOR

1. Evaluate and determine what stock structure is most appropriate for the bigeye tuna stock assessment with consideration of a Pacific wide assessment.

The stock assessment for bigeye tuna in the WCPO is based on the assumption that the western Pacific contains a single stock of bigeye tuna, and hence that there is no [substantial] mixing between the western and eastern Pacific. The area assessed is divided into six interacting regions, with the populations within each region assumed to be perfectly mixed. The Panel was advised that the spatial structure was defined such that:

(i) broad "ecological" regions are reflected;

(ii) there are sufficient fisheries data in each region to enable estimation of important region-specific parameters;

(iii) the regions reflect fishery characteristics, particularly homogeneity of CPUE and size composition as far as possible; and

(iv) for management analyses using the assessments, it is necessary to have the same (or at least comparable) spatial configuration for the three tropical tuna species, skipjack, yellowfin and bigeye.

The Panel noted that the specifications for the regions have changed over time. The assessment assumes that biological parameters such as size-at-age, natural mortality, and maturity / fecundity are constant across the WPO.

The Panel reviewed plots of tag-returns, noting that interpretation of tag-recapture data is best conducted within the framework of a model because a model can better take account of factors such as the distribution of fishing effort, and the growth of tagged animals, and hence their availability to recapture. Nevertheless, the tagging data (tags with unreliable recapture data removed), particularly those for recent years (the PTTP program) suggest that many animals cross the eastern boundary at 150^0W in both directions. Thus, the assumption that the area on which the assessment is based contains a single stock is violated. The estimates of the status of bigeye tuna in WCPO from past Pacific-wide assessments (e.g. Hampton and Maunder 2006) are qualitatively similar to those from the WCPO-only assessment, suggesting that conducting WCPO-only assessments may be a robust approach for the provision of management advice. However, those past Pacific-wide assessments were conducted prior to the availability of the data from the PTTP program. The Panel therefore **recommends** that a Pacific-wide assessment be conducted soon to evaluate whether the past conclusion that the results from a WCPO-only assessment are consistent with expectations from a Pacific-wide assessment remains true. In addition, the Panel **recommends** that Pacific-wide assessments should be conducted regularly (~ every five years) to confirm the assumption that a WCPO-only assessment will provide robust estimates of stock status.

In addition to movement across the 150^0W boundary, there is evidence (e.g., WCPFC-SC7-2011/SA-WP-01) that biological parameters (e.g., maturity-at-length, sex-ratio) for bigeye tuna change from west to east across the WCPO, and that this continues into the EPO. At present, it is not possible to account for this variation in the assessment, but allowing for spatial variation in biological parameters should form a focus for future model development work.

Inferences regarding high rates of movement across the 150^0W boundary are based primarily on the results of a single tagging program, and rates of movement may change over time in response to environmental conditions. The Panel thus **supports continuing** tagging programs to allow estimates of movement rates to be obtained for a wide range of environmental conditions.

The Panel has no specific recommendations in relation to high volume fisheries which catch many small fish, but for which data are uncertain (such as those in Indonesia and the Philippines). However, it **recommends** that it is appropriate to include these fisheries in the assessment to ensure their catches are removed from the population correctly with respect to length. However, the data for these fisheries should not have a large impact on estimates of population trend and size. For example, the model could be fitted to tag-recaptures aggregated across several fleets to avoid it being driven by small sample sizes.

The tag-recapture data for region 3 suggest that the assumption of homogeneity within this region is likely to be violated. The Panel **recommends** considering splitting this region into two (as has been done in the past, e.g., Hampton and Maunder 2006). In addition, the Panel **recommends** examining whether region 5 should be split into two regions to better account for tagging off eastern Australia.

2. Comment on the adequacy and appropriateness of data sources for stock assessment. Evaluate the use (robustness) of modified data from sampling bias studies. Identify data uncertainties and its effects on assessments results. Recommend methods to resolve data uncertainties.

Purse seine length frequency data

The new method to estimate purse seine catch and length frequency is more accurate because it adjusts for the bias observed in the grab sampling and avoids reliance on logsheet estimates, which were found to be biased (higher proportion of skipjack) based on observer data. The new method increased the purse seine catch and generally reduced the sizes in the length frequencies. Using the catch and length-frequency data from the new method in the reference model increases the biomass since about 1980 compared to the old method (Run4_sbest in WCPFC-SC7-SA-WP-02). The influence of the change to the purse seine length-frequency data was shown during the review to be due to the loss of some years (Request M) rather than the bias correction itself. Earlier length-frequency data were collected using port sampling, and may not have the same bias as the grab sampling data. Therefore, it might be possible to use these data in the stock assessment. The sorted and transhipment data may need to be dealt with differently. Most of the early purse seine length-frequency data were from the US fleet, and it is not clear how representative these data are of the whole purse seine fleet.

Request D suggests that re-weighting the data by catch rather than summing the length frequencies leads to length frequencies with smaller fish. The Panel **recommends** that further exploration of methods for weighting purse seine length frequencies by catch be explored. The Panel noted that uncertainty due to length frequency sampling has been modelled for eastern Bering Sea pollock. Request M suggests that the approach used for constructing purse seine length frequencies had minimal impact on biomass estimates, although this conclusion has yet to be confirmed for a broad set of model specifications, in particular when some of the tagging data are omitted from the assessment.

Weighting of Japanese length-composition data

The Japanese longline composition data were weighted by the long-term average CPUE (relative abundance) to correct for spatial variability in the fishery. However, the analysts advised that it may be preferable to weight by long-term catch, to focus the composition data on the catch rather than the population abundance, and to remove fish of the correct size from the population. In addition, there may be little composition data for areas with moderate CPUE. The Panel **supports** weighting the composition data by long-term catch because the composition data would more likely to be collected where catch is taken and there should be less moderately-weighted strata with no composition data. These adjustments are used to account for changing selectivity, and a better approach would be to explicitly model the changing selectivity.

Use of the Japanese longline length-frequency data

There are inconsistencies between some Japanese longline length and weight frequency data. This is particularly apparent in regions 2 and 3 (Fig. 20 of document WCPFC-SC7-2011/SA- WP-02). For example, the mean weight in region 3 is constant over time while the mean length is declining. The model does not fit the length-frequency data in this region. There are also major discrepancies between the observed average catch weights and the average catch weights inferred from mean catch lengths for some regions (see response to Panel request E; Fig. E.1). Assigning more weight to these length frequency data when fitting the model does not lead to markedly improved fits, and also results in poor fits to other composition data (Request N). The weight-frequency data are collected from landings and are generally considered representative of the catch (apart from errors assigning landings to regions, which should be small). In contrast, much of the length-frequency data for the Japanese longline fishery appears to have been collected on training vessel cruises which may have taken place in areas and times different from when and where the majority of catches were taken (anecdotal evidence suggests this is indeed the case, and the reweighting scheme based on long-term CPUE was designed to overcome this problem). The Panel therefore **recommends** that the length-frequency data for the Japanese longline fishery be omitted from the reference model until these data are better understood and can be shown to be compatible with the weight-frequency data. Specifically, the Panel **recommends** that the analysts gain access to how training vessel trips and any other sampling programs are undertaken, and analyse the available data at the set-by-set level before the length-frequency data are considered for re-inclusion in the assessment. The Panel **recommends** separating the training vessel length-frequency data from the commercial data might allow for a separate "survey" length-frequency series to be included in the model. Such data could possibly be used as an index of relative year-class strength and provide information on growth.

Tagging data

The tagging data are very influential on the estimates of absolute biomass. Several of the reporting rate parameters are estimated at the upper bound of 0.9. This is constraining the upper bound on the biomass estimates, and is also constraining the sensitivity of the model estimates to other model changes. The analyses provided to the Panel showed that dropping the tag releases in region 5 (Run P2, request O) led to a marked increase in biomass (and to a recruitment pattern which is more stable over time). Detailed exploration of the results of the reference model and those for Run P2 revealed that the estimates of the reporting rate for the Australian longline fishery in region 5 equalled the pre-specified maximum values (0.9) in the reference model, but this model was still not able to replicate the time-series of recaptures from two of the tag release groups in region 5 (Fig. 1). The Panel reviewed the distribution of releases and recaptures in region 5, and determined that these releases occurred in a small area off eastern Australia (Hampton and Gunn 1998) which was also where the Australian longline fishery started. The approach used to account for lack of mixing does not adequately deal with this problem. The Panel **recommends** dropping these tagging data unless the model can be re-structured to make the area where the Australian tagging took place in region 5 a separate region. Removing additional tag releases that lead to reporting rates at the upper bound also result in increased biomass (Panel request X). Some estimated reporting rates remain at the upper bound, and further analyses are needed to address this issue. The Panel notes that the assessment assumes zero long-term tag loss and tagging-induced mortality. They **recommend** that available data on tag shedding be examined and used to provide a value for use in the assessment, noting that this may be challenging given the possibility of correlation between tag loss for each tag on double-tagged animals. The Panel notes that initial tag loss/tag-induced mortality is currently assumed to be part of the tag reporting rate. The Panel **recommends** modelling these processes separately.

Operational versus aggregated longline CPUE data

The operational data produce CPUE time series only moderately different from the aggregated data, with more temporal variation and is considered by the Panel to be more appropriate than analyses based on aggregated data. The use of operational CPUE data (Run3c) substantially increases the estimates of biomass. Based on Request K, the increase was due to the use of the operational data and not the loss of years of data. The main increase is in region 3, and the trend in abundance in region 3 is also different. The operational CPUE have a few outliers that are unrealistic, but these do not influence the results (Request L). The CVs calculated for the operational data are probably based on differences from the base (first) year and the Panel **recommends** that a more appropriate method should be used to calculate the CVs (e.g. Francis' canonical method or prediction-based methods). There was no vessel call sign information to identify vessels before 1976 so the vessels were all considered the same in CPUE standardization. However, any unidentified vessels in the later years were also considered the same as these early vessels and this may bias the results. Removing these unidentified vessels from the latter period is **advised**. The Panel **recommends** that analysis of operational data focus on how to identify targeting. They recognize that this could be challenging owing to lack of data on sets for which the target was known. The Panel also **recommends** that future analysis using operational data investigate year-area interactions and the implications of increasing numbers of year-area cells without data. Models will need to be developed to interpolate catch rates for cells without data.

Given difficulties of measuring effort, the Panel **agrees** that purse seine CPUE should be given minimal weight in the assessment.

Age and growth

The growth curve estimated in the model overestimates the size of the old fish compared to tagging growth increment data. The tagging growth increment data are from one region and growth there may differ from the other regions in the model. Additional otolith data are available from region 5 that appears to have different growth rates. The estimated growth rates differ if MFCL is run for each region separately. A more comprehensive analysis of the growth information is needed. The Panel **recommends** using methods that simultaneously use both age-length and growth increment data, ideally integrated within MFCL. The Panel notes that the availability of regular collection of age composition data could also improve the estimates of year-class strength and age-specific selectivity.

3. Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

The assessment is based on MFCL, an assessment tool which has been largely developed to assess tuna stocks. The method is appropriate for conducting the assessment of bigeye tuna, although the robustness of the assessment could have been addressed more thoroughly had alternative assessment methods with different model structure options also been applied (e.g., SS3, CASAL). The time to conduct assessments is large, but the analysts had sufficient computing resources that this did not severely constrain the ability to the Panel to explore model variants during the review meeting. The Panel has several recommendations specific to MFCL that are provided in Section 9.

The Panel found that availability of the "viewer" useful to rapidly compare model runs, and evaluate models fits, and encourages its further development and use. The Panel notes that some of the outputs in the viewer could have been clearer. For example, the labelling of the tag-recapture is misleading (it reads "tag group" when the meaning is actually "by fishery group"—e.g., associated and unassociated purse seine fisheries—not "by tag-release"). Also, the overall summaries of the length- and weight-frequencies should be based on weighting the quarter-specific data by the weight assigned when conducting the assessment and not the raw sample size. The Panel **encourages** further developments of this very useful tool. The additional outputs provided in R (e.g. graphs of mean and variation in length and weight composition over time) were also very useful.

4. Evaluate the assessment model configuration, assumptions, input data and configuration, and primary sources of uncertainty, parameters (fishery, life history, and spawner recruit relationships), determine if data is used appropriately, input parameters seem reasonable and primary sources of uncertainty are accounted for.

Reporting rates

The analysts presented the approach to evaluating reporting rates as used in the model. These were based primarily on tag seeding experiments, and included an evaluation of using steelhead tags (to avoid incidental tag loss from fresh placements of conventional tags on dead fish during seeding). They also presented results from a general linear model of factors affecting recapture probability including tagger, release group, smooth (length), species, condition, quality. Data on tag shedding and tag mortality are unavailable for bigeye tuna. The analysts generally assigned reporting rates by flag of vessel, but noted that there may be reporting issues related to unloading location (and perhaps processing method). They fitted a time trend model to reporting rates from seeding experiments which suggested that reporting rates may vary over time, which further complicates use of tag data within the model. The Panel **recommends** continuing seeding experiments due to the impact that reporting rates have on the present model configuration and estimation.

Life history issues

Length data for bigeye tuna are available and suggest a cline in mean length, with larger fish towards the east compared to the west. Information on maturity-at-length using histology was also presented based on a pilot project (WCPFC-SC7-2011/SA-WP-01). Results appeared largely consistent with previous analysis. The Panel noted that maturity-at-age was pre-specified in the assessment and **recommends** that MFCL be modified so that when the maturity data are based on length, that converting to ages is done internally to the model. Presently, the maturity-at-age is based on a fixed age-length relationship. The Panel also **recommends** that the model be modified so that direct length-at-age data can be fitted to assist in estimating growth. With regard to sexual dimorphisms for bigeye, the Panel acknowledged that while the approach for approximating the age-specific vector of natural mortality was appropriate, a gender-specific model would be preferred.

Stock-recruitment relationship
Document WCPFC-SC7-SA-IP-08 outlined the use of meta-analysis to summarize information on steepness for tuna stocks. The Panel **agrees** that this is a useful approach, but notes that the results are currently incomplete because data for several stocks are not included in the analysis and no account was taken of the relative reliability of the assessments that were included. The reference model is currently based on a steepness of 0.8, with sensitivity explored to values of 0.65 and 0.95. The Panel **agrees** that this is an appropriate range for use in the assessment. The impact of the value assumed for steepness is primarily its impact on the results of projections, yield analyses and assessments of stock status in relation to MSY-based reference points. The Panel **recommends** that sensitivity continue to be shown to the assumed value for steepness, and an appropriate means (e.g. a decision table) be used to summarize the management implications of uncertainty regarding steepness.

There is little information regarding the form of the stock-recruitment relationship for bigeye tuna in the WCPO. The Panel did not see value in considering alternative functional forms at present. However, there would be value in exploring a stock-recruitment relationship which determines the average annual recruitment and pro-rates that recruitment to season using parameters which are constant over time instead using of a stock-recruitment relationship which assumes that the average recruitment for a given level of spawning biomass is independent of season. Seasonal patterns in recruitment lead to additional recruitment variability, which may be dampened by applying the same stock-recruitment penalty (recruitment deviate penalty) to all seasons.

The final biomass for the reference model and Run 21 (in WCPFC-SC7-SA-WP-02) differ. This is unexpected because the only difference between these two runs is to which recruitments the stock-recruitment relationship applies. Exploration (Request AB) indicated that the size of the penalty imposed on the stock-recruitment deviations impacts the level of the biomass. The Panel **recommends** that the size of the penalty be selected which allows the asymptote of the stock-recruitment relationship to be estimated, but is otherwise uninformative about historical stock size. The results for Request AB suggest that a standard deviation of the log-residuals of 1.5 and larger is sufficient for the stock-recruitment penalty to be inconsequential in terms of its impact on biomass. The Panel also **recommends** consideration of fitting the stock-recruitment relationship to the annual rather than seasonal recruitments. Finally, the Panel notes that projections should be based on the estimate of the standard deviation of the log-residuals about the fitted stock-recruitment relationship and not the pre-specified value for this standard deviation.

Selectivity assumptions
Currently, the Chinese Taipei fishery is the only one for which selectivity is assumed to be asymptotic. This choice was made primarily because this fleet catches the largest bigeye. Request AD evaluated the sensitivity to changing the assumption about Japanese longline selectivity. Results showed that the biomass was smaller if selectivity for the Japanese longline fishery was assumed to be asymptotic. This highlights the importance of carefully evaluating the choice and constraints of selectivity patterns and understanding the mechanisms for dome-shaped selectivity.

Data weighting
The Panel reviewed the scheme used in the reference model for weighting the various data sources (see Table E.1 for a summary of the weights used in the reference model). Some of the data sets (e.g. the catch length- and weight-frequency data for the Chinese Taipei longline fishery) are severely downweighted in this model. However, the weights for most of other data sets were assumed to be the same (and relatively high; 20 in Table E.1). The Panel **recommends** that the statistical weights for each data set be re-evaluated and revisited with each subsequent assessment. A preliminary re-weighting was undertaken during the review (see Request W).

6. Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results.

The assessment reflects uncertainty in several ways, including sensitivity tests, asymptotic variance estimates for key model outputs, and Kobe plots. These approaches are state-of-the-art in fisheries stock assessment. The only standard measure of uncertainty not included in the assessment is a retrospective analysis (both a 'standard' retrospective analysis where the same assessment set-up is applied to data sets that have been truncated and a 'historical' retrospective analysis where the results of recent assessments are contrasted). The Panel **recommends** that future assessments include both standard and historical retrospective analyses.

The Panel found the results of "structural analysis" (e.g. Fig. 58 of document WCPFC-SC7-2011/SA-WP-02) to be a particularly successful way to convey uncertainty. Structural analysis involves applying the assessment method to a crosswise grid of many combinations of assumptions. The value of this type of summary output will be enhanced if probabilities can be assigned to each factor included in the structural analysis and these probabilities accounted for in the graphical summary (e.g. by making the size of each circle proportional its probability or by creating a probability surface). The Panel therefore **recommends** that methods be developed to provide output which accounts for uncertainty regarding the values for the factors considered in the structural analysis. The Panel cautions that assigning probabilities to factors can be extremely challenging, and appropriate resources should be assigned by the WCPFC SC to this task, were it decided to assign probabilities to each factor.

7. Comment on the proposed reference points and management parameters (e.g., MSY, F_{msy}, B_{msy}, MSST, MFMT); if possible and feasible, estimate values for alternative reference points (or appropriate proxies) and view on stock status.

The Panel notes that the assessment reports projection results relative to standard management reference points (e.g. F_{MSY} and B_{MSY}) (e.g., WCPFC-TCC7-2011-31). The Panel **agrees** that given that steepness is pre-specified in the assessment, estimation of F_{MSY} is more robust than estimating B_{MSY} which depends on both steepness and the asymptote of the stock-recruitment relationship. Estimation of the latter for bigeye tuna in the WCPO depends on whether the recent or entire sequence of recruitment estimates forms the basis for fitting the stock-recruitment relationship. The Panel notes that recent recruitment would be appropriate if (a) expected recruitment has changed due to a "regime-shift" and MSY pertains to "prevailing environmental conditions" or (b) the early recruitments are deemed biased. In contrast, the Panel **agrees** that, *a priori*, stock-recruitment relationships should be based on the entire period for which estimates of spawning biomass and recruitment are available so that the resulting stock-recruitment is an average over environmental regimes. If data from the earlier period of the model are less reliable (i.e., have higher variance specified on input) then the fitting of the stock-recruitment relationship within the model should account for variable uncertainty over the model period. The estimates of recruitment for recent years are more robust to changes in model specifications. The Panel examined plots of recruitment versus time, recruitment versus catch, and recruits/spawner versus time and spawning biomass based on the output from the reference model. However, none of the information provided to the Panel enables a definitive conclusion to be drawn between the two options for which data should form the basis for estimating the stock-recruitment relationship.

The Panel notes that fishing mortality relative to F_{MSY} is generally more robustly estimated than stock status relative to B_{MSY} given a pre-specified value for steepness (but note the discussion regarding how the stock-recruitment relationship is fitted under ToR 4). In addition, constant fishing mortality rate strategies have been shown to be robust to regime-shift like changes in recruitment. Consequently, management based solely on a constant fishing mortality strategy based on a fishing mortality of F_{MSY} (or less) would lead to robust management advice (the catch recommendation would be the same irrespective of which recruitment assumption is true) and would in time lead to recovery to B_{MSY} (or greater) if F_{MSY} and current biomass are estimated accurately. However, the time to recover to B_{MSY} would depend on which recruitment assumption is true. Times to recovery and the extent to which other management objectives are satisfied would therefore need to be reported regularly.

The Panel was presented with the method used to bias-correct the stock-recruitment relationship when calculating management reference points. The Panel **agrees** that bias-correction be applied if estimates from the deterministic yield function are presented. However, it also **recommends** that stochastic yield functions be presented because they may not indicate the same values for management reference points such as F_{MSY} and B_{MSY}.

8. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status. This would include the methods of projection under hypothetical various options in future management measures (on effort? On catch? By fisheries? Etc.)

The approaches used to conduct future projections are fairly standard. The Panel was satisfied that the approaches taken are appropriate. Discussions related to projections and projection methodology are outlined under ToR 4 and ToR 7. The Panel noted that MSY depends on the mix of fisheries. The Panel **recommends** that such variations be considered in projections should fishery-specific management measures be evaluated.

9. Suggest research priorities to improve our understanding of essential population and fishery dynamics, necessary to formulate best management practices.

The Panel identified a number of research activities and general methodological recommendations which, if addressed, should improve the ability of the assessment to provide scientific advice for management decision making. The Panel notes that its ability to fully evaluate the assessment was hampered by the lack of availability for analysis of some of the raw data. In particular, the Panel wished to understand whether the trend in operational CPUE for the early years of the assessment period (to which the results have been shown to be sensitive) were due to a lack of vessel codes for some vessels (which may have fished into the years for which vessel codes are available), as well as details of the length-frequencies which are collected on Japanese training vessels. However, the lack of raw data precluded these tasks.

Key recommendations from this review include:

1) When moving from one reference model to a modified one, care should be taken to change only one factor at a time to ensure the impact of changes can be fully understood.
2) The way the fisheries are linked should be more fully documented in the assessment report, and the implications of such linkage should be more fully evaluated.
3) A Pacific-wide assessment should be conducted soon to evaluate whether the past conclusion that the results from a WCPO-only assessment are consistent with expectations from a Pacific-wide assessment remains true.
4) Pacific-wide assessments should be conducted regularly (~ every five years) to confirm the assumption that a WCPO-only assessment will provide robust estimates of stock status.
5) Continue tagging programs to allow estimates of movement rates to be obtained for a wide range of environmental conditions
6) High volume small-fish fisheries (e.g., Philippines and Indonesia) should be retained in the model to ensure their catches are removed from the population correctly with respect to length. However, the model should be formulated so that the data for such fisheries do not have a large impact on estimates of population trend and size.
7) Consider splitting regions 3 and 5 each into two regions.
8) Further explore methods for weighting purse seine length frequencies by catch.
9) Further explore methods for the calculating longline size-composition data by weighting spatial data by long-term average catches.
10) Length-frequency data for the Japanese longline fishery should be omitted from the reference model until these data are better understood and can be shown to be compatible with the associated weight-frequency data. Analysts should gain access to how training vessel trips

and any other sampling programs are undertaken, and analyze the available data at the set-by-set level before these length-frequency data are considered for re-inclusion in the assessment.

11) Separate the training vessel length frequency data from the commercial data and create a "survey" length composition series to be included in the model.
12) A more appropriate method should be used to calculate the CVs for the Japanese CPUE indices (e.g. Francis' canonical method or prediction-based methods)
13) Drop the region 5 tagging data unless the model can be re-structured to make the area where the Australian tagging took place in region 5 a separate region.
14) Available data on tag shedding should be examined and be used to provide a value for use in the assessment, noting that this may be challenging given the possibility of correlation between tag loss for each tag for double-tagged animals.
15) Tag loss and tagging-induced mortality should be modeled separately.
16) Future analysis of operational CPUE data should focus on how to identify targeting and investigate year-area interactions and the implications of increasing numbers of year-area cells without data.
17) Use methods that simultaneously use both age-length and growth increment data, ideally within MFCL.
18) Continue seeding experiments due to the impact that reporting rates have on the present model configuration and estimation.
19) Sensitivity analyses should continue to be shown to the assumed value for steepness and an appropriate means (e.g., a decision table) used to summarize the management implications of uncertainty regarding steepness.
20) The size of the stock recruitment penalty should be selected which allows the asymptote of the stock-recruitment relationship to be estimated, but is otherwise uninformative about stock size.
21) Consider fitting the stock-recruitment relationship to the annual rather than seasonal recruitments.
22) The statistical weights for each data component (e.g., size composition, tagging, effort deviations) should be re-evaluated and revisited with each subsequent assessment.
23) Future assessments should include both standard and historical retrospective analyses.
24) Methods should be developed to provide output which accounts for uncertainty regarding the values for the factors considered in the structural analysis.
25) Stochastic yield functions should be presented because they may not indicate the same values for management reference points such as F_{MSY} and B_{MSY}.
26) Projections considering MSY estimates should account for fishery-specific changes (i.e., likely proportional catches by fishery).
27) The following recommendations relate to MFCL:
    a. Test the options for time-varying selectivity – allowing for time-varying selectivity may address some of the issues related to the sometimes poor fits to the length- and weight-frequency data.
    b. Allow the length bins to be of different widths. One might, for example, want many narrow length bins for the smaller lengths, but fewer but wider length bins for the larger lengths. Allowing for a more flexible length bin structure should also reduce computational times as well as better reflect the available data.
    c. Allow for long-term and initial tag-loss. Currently initial tag-loss is implemented by reducing the number of animals tagged when inputting data to the model and no account can be taken of long-term tag-loss.
    d. Include an option which allows the tagging data to inform movement only rather than movement and mortality.
    e. Allow conditional age-at-length data to be included in the likelihood function. This will allow the ageing data from current sampling (e.g. WCPFC-SC6-2010/GN IP-04) to be formally included in the assessment.
    f. Extend MFCL to allow gender to be explicitly represented. This will allow the impacts of differences in growth and natural mortality between the sexes to be represented. The current approach to modeling, for example, length-specific natural

mortality (e.g. WCPFC-SC4-2008/ ME-WP-1) seems unnecessarily complicated given the lack of gender-structure in the model.

g. Create an output table which lists all of the likelihood components by fleet and automates the process of computing effective samples sizes (and other summary statistics related to model fit).
h. Allow for more general selectivity options, including selectivity patterns where the first age for which selectivity is non-zero is pre-specified. This should help to avoid selectivity being non-zero owing to the functional form for selectivity rather than data.
i. Include a “tail compression” option, which would pool all length- and weight-data for large and small sizes based on a specified percentage (e.g. all lengths would be pooled so that the “plus” length-class contained 0.1% of the length-frequency).
j. Add an option which allows the analyst to assume a multinomial likelihood for the compositional data in the first phases and only transition to the robust normal likelihood in the later phases.
k. When maturity data are based on length, converting to ages should be done within the model. Presently, the maturity-at-age is based on a fixed age-length relationship.
l. An option to add a likelihood weight to the tagging data component should be added.

References

Hampton, J. Andand Gunn, J.S. (1998) Exploitation and movements of yellowfin tuna (*Thunnus albacares*) and bigeye tuna (*T. obesus*) tagged in the north-western Coral Sea. *Mar. FreshwaterFreshw. Res*. 49: 475-489.

Hampton, J. and Maunder M.N. (2006) An update of Pacific-wide assessment of bigeye tuna with comparisons with eastern Pacific assessment results. IATTC Working Group on Stock Assessments, 6th Meeting, SAR-7-07c.ii.

Figures

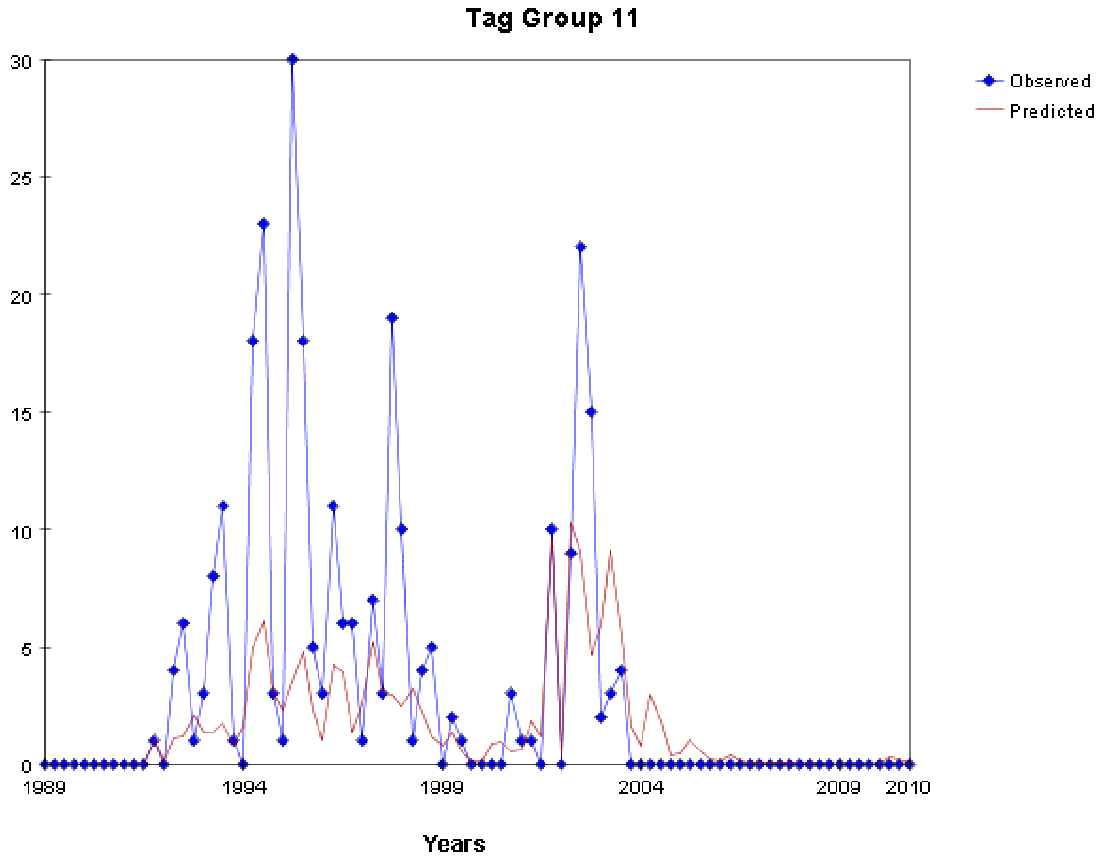


Figure 1. Observed and model predicted tag returns from region 5 (tag group 11) of bigeye tuna over time.

Appendix A: Panel Biographies

James Ianelli is an affiliate professor at the University of Washington and a stock assessment scientist with the Resource Ecology and Fisheries Management division of the Alaska Fisheries Science Center. He earned his PhD in Fisheries Science at the University of Washington in 1993, and is part of the stock analysts responsible for producing annual assessments of a number of important groundfish species in the North Pacific. His research interests include developing statistical approaches for ecosystem/fisheries conservation management. He chairs the North Pacific Fishery Management Council's groundfish Plan Team and serves on the Advisory Panel for the Commission for the Conservation of Southern Bluefin Tuna. He is a senior editor for the journal *Natural Resource Modeling*. He has conducted annual stock assessments for a number of species in Alaska starting in 1992. He serves as an advisor to the South Pacific Regional Fisheries Management Organization involved with assessments of Chilean jack mackerel.

Mark Maunder is the Head of the Stock Assessment Program at the Inter-American Tropical Tuna Commission. He received his B.Sc (Zoology and Computer Science), M.Sc (Zoology) at the University of Auckland and Ph.D. (Fisheries) at the University of Washington. Before joining the IATTC , Dr Maunder was a Quantitative Fisheries Scientist at the New Zealand Fishing Industry Board. His research interests include development of statistical methodology for fisheries stock assessment, protected species, and ecological modeling. He has published over 50 papers in the peer-reviewed literature, along with many technical reports. Dr Maunder is co-founder and president of the AD Model Builder Foundation and a member of the Partnership for Mid-Atlantic Fisheries Science (PMAFS) Science Advisory Committee.

André E. Punt is a Professor of Aquatic and Fishery Sciences at the University of Washington. He received his B.Sc, M.Sc and Ph.D. in Applied Mathematics at the University of Cape Town. Before joining the University of Washington, Dr Punt was a Principal Research Scientist with the CSIRO Division of Marine and Atmospheric Research. His research interests include the development and application of fisheries stock assessment techniques, bioeconomic modelling, and the evaluation of the performance of stock assessment methods and harvest control rules using the Management Strategy Evaluation approach. He has published over 190 papers in the peer-reviewed literature, along with over 400 technical reports. Dr Punt is currently a member of the Scientific and Statistical Committee of the Pacific Fishery Management Council, the Crab PLAN Team of the North Pacific Fishery Management Council, and the Scientific Committee of the International Whaling Commission. He is the Associate Editor of the journals *Fisheries Research*, *Population Ecology*, and the *Journal of Applied Ecology*.

Appendix B: Final TOR

WESTERN AND CENTRAL PACIFIC FISHERIES COMMISSION

TERMS OF REFERENCE

PEER REVIEW OF THE 2011 BIGEYE TUNA STOCK ASSESSMENT

Introduction

Based on the WCPFC's consultancy *Independent Review of the Commission's Transitional Science Structure and Functions*, the Fifth Regular Session of the Commission (WCPFC5) in 2008 requested the Secretariat prepare a proposal that would support the periodic peer review of ISC and SC stock assessments. The Fifth Session of the Scientific Committee (SC5) prepared a general process of the peer review and WCPFC6 agreed that the latest full bigeye stock assessment for the WCPO be reviewed. According to the recommendation from SC6 that the Commission allocate a budget to carry out the bigeye stock assessment peer review, WCPFC7 endorsed undertaking a workshop-style peer review of the 2011 bigeye stock assessment in early 201.

SC7 developed the process for the peer review of the 2011 bigeye tuna stock assessment, including formulation of the peer review panel, selection procedure and time frame, budget, and terms of reference for the review. This project is based on the process prepared by SC7.

OBJECTIVE OF THE ASSIGNMENT

Using Articles 13 Paragraph 4 of the Convention as a basis, understanding the recommendations from the *Independent Review of the Commission's Transitional Science Structure and Functions*, and following the process agreed at SC7, undertake, in consultation with the stock assessment team (SPC), a peer review of 2011 bigeye tuna stock assessment in the western and central Pacific Ocean (WCPO).

SCOPE AND TASKS

The Review Panel will conduct the following tasks, but not limited to, and provide comments recommendations, if needed, upon issues additional to those listed below:

1. Evaluate and determine what stock structure is most appropriate for the bigeye tuna stock assessment with consideration of a Pacific wide assessment.
2. Comment on the adequacy and appropriateness of data sources for stock assessment. Evaluate the use (robustness) of modified data from sampling bias studies. Identify data uncertainties and its effects on assessments results. Recommend methods to resolve data uncertainties.
3. Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.
4. Evaluate the assessment model configuration, assumptions, input data and configuration, and primary sources of uncertainty, parameters (fishery, life history, and spawner recruit relationships), determine if data is used appropriately, input parameters seem reasonable and primary sources of uncertainty are accounted for.
5. Particular attention is to be paid to the following;
    - A) Length of older individuals and the impact it has on the stock assessment results.
    - B) Potential for regime shift in recruitment. Consider whether shifts in recruitment are real or are caused by model artifacts.
    - C) Appropriateness of the stock recruitment relationship.
    - D) Availability of bigeye to purse seine and not being available to longline.
    - E) Investigate the cause of residual patterns in the length composition data and determine how it can be resolved.

F) The use of CPUE indices in the assessment (purse seine, pole-and-line and/or longline) and consider the regional weighting of these standardized indices.
G) Determine if the manner in which the movement and tagging data are modeled is appropriate.
H) Determine if the spatial structure of the model is appropriate.

6. Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results.
7. Comment on the proposed reference points and management parameters (*e.g.*, *MSY*, *Fmsy*, *Bmsy*, *MSST*, *MFMT*); if possible and feasible, estimate values for alternative reference points (or appropriate proxies) and view on stock status.
8. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status. This would include the methods of projection under hypothetical various options in future management measures (e.g. On effort? On catch? By fisheries? Etc.)
9. Suggest research priorities to improve our understanding of essential population and fishery dynamics, necessary to formulate best management practices.

OUTPUTS AND SCHEDULE

Task	Timeframe
Review of the 2011 WCPO bigeye stock assessment results (possibly including all the input data, modeling software, output of basic runs as well as all the sensitivity runs)	Prior to the Peer Review Workshop
Reviewers participate in the review workshop at SPC, Noumea, New Caledonia (approximately 2 days on peer review of the 2011 assessment and a further 3 days on reviewing and advising on various aspects of subsequent assessments).	2 - 6 April 2012
Send a draft review report to SPC for review and response	Within one week after the workshop
Submit final review report and SPC's response to the WCPFC Executive Director	On or before 31 May 2012

Annotated version of ToR

1. Evaluate and determine what stock structure is most appropriate for the BET stock assessment with consideration of a Pacific wide assessment.

- The current eastern boundary is primarily political
    - What impact might it have on the management advice from the assessment and how might we evaluate it?
    - What biological assumptions are reasonable and can they be approximated / modelled?
- Our regional assessment structure is rather pragmatic seeking to balance patterns in the data, seeking consistency across stocks, and lining up with management needs. However, we see strong differences in longline size composition data across regions and some evidence that at least growth rates differ.
    - What are some principles to determine appropriate regional structure?
    - How might we examine alternative hypotheses for the patterns in longline size data using the existing assessment model including the impacts of alternative assumptions about maximum size?
- The assessment assumes complete mixing of fish within regions yet there does not seem to be the expected response in longline CPUE from increased purse seine catches in region 3.
    - Does this represent some potential biological process not being captured, or is it simply a model artefact?
- We have large fisheries in Indonesia and the Philippines – many of which catch very small fish. These data are highly uncertain.
    - What are some ways to model the impact of these fisheries on the overall assessment?

2. Comment on the adequacy and appropriateness of data sources for stock assessment. Evaluate the use (robustness) of modified data from sampling bias studies. Identify data uncertainties and its effects on assessments results. Recommend methods to resolve data uncertainties.

- <u>Purse seine catches</u>: our assessment have included two alternative estimates of purse seine catches and size composition data. The 'newer' estimates are based on an active research program and <u>will be subject to an independent review in the second half of 2012</u>.
    - What is the relative plausibility of the two sets of estimates?
    - Noting that we have a review upcoming, do you have any thoughts on the appropriateness of our methods to correct the purse seine catch and size composition data?
- <u>Size composition data</u>: our models are very sensitive to trends in size data and our fleets move around (within an region) and catch different sized fish.
    - How best can compensate for this shifting fleets and fish sizes to give the model consistent data that also allows for appropriate removal of catches? Does this explain the phase shifts observed in the residuals for some of the length data sets?
    - How much value could regularly collected direct ageing data add to our assessments?
- <u>Tagging data</u>: tagging is occurring within parts of specific regions within our assessment and we now have several years of continuous release data. Model estimates of biomass are extremely sensitive to the influence of tagging data.
    - How should we best model localised releases?
    - How should we model recaptures outside (often just outside) the model domain?
    - How valuable to the assessments could be lower levels of annual tagging?
    - How important is it to continue research into reporting rates during the tagging programme? If important, how long after tagging finishes should we continue this work?
    - Should we allow the tagging data to influence our estimation of growth?
- <u>Longline CPUE data</u>: the assessment draws heavily on the Japanese longline CPUE as this fleet has been in operation since the beginning of the fishery. These data are critically influential. In recent years there has been considerable reduction in the effort of this fleet and

areas fished and new fleets are becoming more dominant. Recently we have had access to Japanese operational data under a collaborative arrangement and are using these indices.

- Has the operational data lead to an improvement to the quality of the assessment?
- Do we have sufficient operation data to appropriately standardize the effort for changes to catchability (targeting etc.)? If not, how can we examine this issue within the assessment?
- How might the catch of other species be used in CPUE analyses?
- How should we handle the reduced effort of the Japanese fleet and the increasing concentration of their effort? Is there the potential for hyperstability?
- Is there value in standardizing CPUE for other longline fleets – would the availability of operational level data be critical to make it worthwhile?

- Purse seine CPUE data: we wish for these data to have no influence on the assessment – especially given their uncertainty.
  - Are we right to reduce their influence? If yes, have we done it correctly?

3. Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

- Assessment method: MULTIFAN-CL is the primary method for assessing tuna stocks in the WCPO though we have undertaken several 'parallel' assessments using Stock Synthesis in the past.
  - Are there any major weaknesses in the MULTIFAN-CL approach given the information available?
  - How often should parallel assessments be undertaken and what alternative approaches might be considered?

4. Evaluate the assessment model configuration, assumptions, input data and configuration, and primary sources of uncertainty, parameters (fishery, life history, and spawner recruit relationships), determine if data is used appropriately, input parameters seem reasonable and primary sources of uncertainty are accounted for.

- Regional weighting: We use the Japanese longline CPUE over the model domain over a period when fishing was widespread to estimate the proportion of the overall stock biomass in different regions over a given period. This is implemented in the model via re-scaled CPUE (and consequently 'effective effort') and shared catchability parameters. We already assume that longline selectivity is not constant among some regions and likely neither is catchability.
  - How appropriate is the current approach? and what alternatives should be considered either for the reference model or for sensitivity analyses?
- Natural mortality: Is currently estimated outside the model and then fixed within (with alternatives compared for sensitivity). M can be estimated from fitting the model with high relative weight assigned to the tagging data and other data greatly down weighted.
  - Is this a valid approach to determine either priors for full estimation of M or value for sensitivity analyses?
- Stock recruitment relationship: We assume a Beverton Holt SRR and provide results across a range of steepness values. Our 'reference case' assumed steepness of 0.8 based on some preliminary analysis of several tuna stocks.
  - Are other functional forms or assumptions around steepness more appropriate?
  - Should we include the lognormal bias correction (more relevant for sections of reference points and projections)?
- Recruitment trend: there is a large trend in overall recruitment driven by region 3. Many things contribute to this, but the conflict between increasing purse seine catches and longline CPUE seems to be a strong driver.
  - Real regime shift or model artefact? If an artefact, is the early recruitment too low or the later recruitment too high?
  - Are we modelling the longline CPUE data series appropriately? Is the first half wrong or the second half of the series providing a trend not reflective of abundance?
- Data weighting: this is an important issue in most assessments. We typically examine sensitivity of assessment results to alternative weighting assumptions.

- Are appropriate weights give to the input datasets? Should a different approach be applied to determining the data weighting?

5. Particular attention is to be paid to the following; (note: we removed these and placed them where appropriate under other ToRs) included in the specific issues mention in 1 – 4 above).

6. Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results.

- Uncertainty estimation: we undertake a full cross grid of the 'one-change' sensitivity analyses and the primary results are based on a reference case and 'plausible' one-change model runs.
  - Should we incorporate the results from the grid into the estimation of uncertainty? If so, how should we do it, e.g., weighting of individual model runs?

7. Comment on the proposed reference points and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if possible and feasible, estimate values for alternative reference points (or appropriate proxies) and view on stock status.

- Lognormal bias correction:
  - Should this be used in the calculation of MSY and related quantities?
- Temporal trend in recruitment:
  - If it is real or an artefact we cannot fix, how best do we estimate stock status and yields?

8. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status. This would include the methods of projection under hypothetical various options in future management measures (on effort? On catch? By fisheries? Etc.)

See discussion on uncertainty and the lognormal bias correct above.

9. Suggest research priorities to improve our understanding of essential population and fishery dynamics, necessary to formulate best management practices.

- Are further biological or tagging studies to examine stock structure warranted?
- How important is further Pacific-wide modelling work?
- Is there a case to split region 3 into 2 to create a separate region for Indonesian and Philippines fisheries?
- Do the spatial differences is sizes observed in the longline fisheries warrant further examination. If yes, should we focus on biological studies, modelling exercises, or both?
- Is further investigation of purse seine catch and size data warranted? If yes, is the current research approach appropriate or should we be trying other approaches?
- Should we continue to develop the Japanese operational longline data? If yes, what emphasis should be put on the early period and 'data rescue' efforts?
- How important is further investigation of regional weightings? If important, should the focus be on operational Japanese data of other information sources / approaches?
- Given the low likelihood that the true value of steepness (if it exists) can be found, is there still value in further research in this area? If yes, should it focus on meta-analyses or modelling from biological first principles?

Appendix C: Background documents considered by the Panel

Davies, N. et al. 2011. Stock assessment of bigeye tuna in the western and central Pacific Ocean. WCPFC-SC7-SA-WP-02.

Davies, N. et al. 2011. Update of recent developments in MULTIFAN-CL and related software for stock assessment. WCPFC-SC7-SA-IP-04.

Hampton, J. and P. Williams. 2005. A description of tag-recapture data for bigeye tuna (Thunnus obesus) in the western and central Pacific Ocean. Col. Vol. Sci. pap. ICCAT 57(2): 85-93.

Harley, S.J. 2011. Preliminary examination of steepness in tunas based on stock assessment results. WCPFC-SC7-SA-IP-08.

Harley, S. J., and N. Davies. 2011. Evaluation of stock status of bigeye, skipjack, and yellowfin tunas against potential limit reference points. WCPFC-SC7-MI-WP-04.

Harley, S.J. et al. 2010. Background information for the 2010 bigeye assessment. WCPFC-SC6-SA-WP-01.

Hoyle, S. 2011. Tag reporting rate prior distributions for the 2011 bigeye, yellowfin, and skipjack stock assessments. WCPFC-SC7-SA-IP-10.

Hoyle, S., and A. Langley. 2011. Spatial size data stratification for length-based stock assessments. WCPFC-SC7-SA-IP-09

Hoyle, S., and S. Nicol. 2008. Sensitivity of bigeye stock assessment to alternative biological and reproductive assumptions. WCPFC-SC4-2008/ ME-WP-1

Hoyle, S., and H. Okamoto. 2011. Analysis of Japanese longline operational catch and effort for bigeye tuna in the WCPO. WCPFC-SC7-SA-IP-01.

Hoyle, S. et al. 2010. Analysis of Japanese longline operational catch and effort for bigeye tuna in the WCPO. WCPFC-SC6-SA-WP-02.

Kleiber, P., et al. 2012. MULTIFAN-CL User's Guide February 2012.

Lawson, T. 2011. Purse-Seine length frequencies corrected for selectivity bias in grab samples collected by observers. WCPFC-SC7-ST-IP-02.

Lawson, T. 2008. Factors affecting the use of species composition data collected by observers and port samplers from purse seiners in the western and central Pacific Ocean. WCPFC-SC4-2008/ST-WP-3.

Lawson, T., and P. Sharples. 2011. Report on Project 60: Collection and evaluation of purse seine species composition data. WCPFC-SC7-ST-WP-03.

Nicol, S., et al. 2011. Bigeye tuna age, growth and reproductive biology (project 35). WCPFC-SC7-2011/SA-WP-01.

Nicol, S., et al. 2010. Pacific tuna tagging project progress report and workplan for 2010. WCPFC-SC6-2010/GN IP-04.

OFP. 2011. Projections based on 2011 assessments (updated analysis with projection results in attached Excel file). WCPFC-TCC7-2011-31

OFP. 2011. Report from the SPC pre-assessment workshop, Noumea, April 2011.

Preece, A., et al. 2011. Identification of candidate limit reference points for the key target species in the WCPFC. WCPFC-SC7-2011/MI-WP-03.

SPC-PFP. 2011. The western and central Pacific tuna fishery: 2010 overview and status of stocks. WCPFC8-2011-IP-02.

Appendix D: Participants during the review

Review Team
Jim Ianelli: NOAA, NMFS
Mark Maunder: IATTC
André Punt: University of Washington

WCFPC
Anthony Beeching

SPC
Aaron Berger
Nick Davies
John Hampton
Simon Hoyle
Shelton Harley
Pierre Kleiber
Dale Kolody
Tim Lawson
Simon Nicol
Peter Williams

Appendix E: Requests to the Analysts

Requests made prior to the meeting

Request A
Request: Impose a multiplicative trend from 1 in 1980 to 0.5 in 2011 to the Japanese longline CPUE (effort) for region 3.
Rationale: The Panel wished to understand whether the increasing trend in recruitment was a result of the increased purse seine catch in conjunction with flat longline CPUE. If this is the case, a declining longline CPUE should eliminate the trend in recruitment.
Response: The biomass estimates were substantially larger than for the reference model. However, the trend in recruitment remained. This suggests that the increasing trend in purse-seine catch combined with the flat CPUE for the longline fisheries in region 3 is not the sole reason for the increasing trend in recruitment.

Request B
Request: Downweight the size and weight data for the longline fishery in region 3 and fix selectivity to that estimated in the reference case.
Rationale: The Panel wished to understand whether the increasing trend in recruitment was a result of the increased purse seine catch in conjunction with flat longline mean weight.
Response: The results were qualitatively the same as those for the reference model. This suggests that flat longline mean length and mean weight in region 3 is not the sole reason for the increasing trend in recruitment.

Request C
Request: Combine the factors considered in requests A and B
Rationale: The Panel wished to understand whether the increasing trend in recruitment was a result of the increased purse seine catch in conjunction with flat longline mean weight and flat longline CPUE.
Response: The biomass estimates were substantially larger than for the reference model, but the trend in recruitment remained. Thus, flat longline mean length and mean weight, combined with flat longline CPUE and increasing purse-seine catch in region 3 is not the sole reason for the increasing trend in recruitment.

Requests made during the meeting

Request D
Request: Compare the length-frequency data for the purse-seine catches currently used in the assessment with those which would be used if the set-by-set data were catch-weighted. Scale to the catch by 5x5 or 1x1 square by quarter. Consider regions 3 and 4, and show results by association type for 1995, 2000, 2005, and 2010**.**
Rationale: The catch length-frequency varies within each region. The current (unweighted) approach is only valid if the length-frequency samples are taken proportional to the catches
Response: The purse-seine length-frequencies were reweighted by 5x5 square, and the resulting length-frequency distributions are moved towards smaller fish. Thus, the construction of catch length-frequencies should be based on catch weighting. However, it was noted that weighting length-frequency samples by catches by 1x1 square may lead to large numbers of 1x1 squares with catch, but no length-frequency data.

Request E
Request*:* Convert the longline mean lengths into mean weights using the length-weight relationship, and plot time-trajectories of observed and predicted mean weight by quarter and fleet.
Rationale: The Panel was concerned that the length and weight data for some of the fleets were inconsistent, perhaps due to differences in the way the data are collected.
Response: There are notable inconsistencies between the observed and mean length-predicted mean weights for some of the Japanese fisheries, but not the Chinese Taipei fishery (e.g., Fig. E.1). This

suggests that Japanese length frequency data are probably not representative of the catch, and should therefore not be used in the assessment until further validation is conducted.

Request F
Request: Compare the likelihood components for the length, weight and tagging data for Runs3J and 19. Construct a function in R so it can be used in comparisons of other runs.
Rationale: The Panel wished to understand how the data component fits are influenced by the changes to the model.
Response: An R routine was written to compare likelihood components for the multiple model runs.

Request G
Request: Determine where the Australian tagging program took place in region 5.
Rationale: The location of tag releases can influence the location of recaptures, and the reference model is unable to mimic the recaptures of these tagging programs.
Response: The tags were released in a spatially-restricted region off Cairns.

Request H
Request: Provide a table summarizing the scaling weights for the length and weight composition data and their average sample size for the reference model.
Rationale: There are a large number of composition data sets included in the assessment and it is difficult to determine which of these are influential in the analysis. In addition, the reasons for the weights were not documented.
Response: The table was presented (Table E.1).

Request I
Request: Provide a table which indicates which fisheries have linked selectivities and catchabilities.
Rationale: Sharing catchability and selectivity among gears can have a large impact on results. The Panel wished to fully understand how the model is specified in this regard.
Response: The table was presented (Table E.1).

Request J
Request: Provide effective sample sizes averaged over all years for the reference model.
Rationale: The effective sample sizes provide a summary of how well the model fits the composition data
Response: The table was presented (Table E.2) using the R code:

```
get.effn <- function (predlf, obslf) {
  effn<-sum((1-predlf)*predlf)/sum((obslf-predlf)^2)
  return(effn)
}
```

Request K
Request: Remove the aggregate longline CPUE for 1950 to 1959 from Run 3b.
Rationale: Two factors changed between Runs 3b and 3c (reduction in the number of years with CPUE data and replacement of the aggregated with the operational CPUE data). The Panel wished to understand which change had the largest impact.
Response: Removing the CPUE data for 1950 to 1959 did not impact the estimates of biomass. The increase in biomass between Runs 3b and 3c is therefore related to the use of the operational rather than the aggregated data.

Request L
Request: Remove the operational longline CPUE data point for 1960 in region 3 from Run 3c (also last CPUE point in region 2)
Rationale: Outliers are unrealistic and may be influential
Response: The run was conducted, but did not lead to a change in results.

Request M
Request: Reduce the number of years with purse-seine length-frequency in Run 3d to match that in Run 3e
Rationale: Two factors changed between Runs 3b and 3c (reduction in the number of years with purse-seine data, and replacement of the grab-sampling-based with the spill-sampling-based length-frequencies). The Panel wished to understand which change had the largest impact.
Response: Changing the number of years with purse-seine length-frequency data has a much larger impact on the estimated biomass trajectory than changing how the length-frequencies are constructed.

Request N
Request: Construct a variant of the reference model (Run 3j) with large weight (1) on the Japanese longline length frequency data for region 3.
Rationale: The model is not fitting these length-frequency data well and the Panel felt that it would be informative to see what in the model would have to change to fit these length frequencies.
Results: This is Run P1. The estimated biomass trajectory is substantially lower than for the reference model. The predicted time-series of mean lengths is similar to that for the reference model, but the fits are somewhat better. The standard deviations of length-at-age are smaller. Recruitment is lower, particularly during the early period of the assessment. Run P1 did not fit the Chinese Taipei and some other composition data very well. Similar results were obtained when the weight frequency data for Japanese longline length composition data for region 3 are less emphasized when fitting the model. It is clear that there is conflict in the data. This run was part of the rationale for the Panel's recommendation to drop all Japanese longline length-frequency data until these data are better understood.

Request O
Request: Construct a variant of the reference model (Run 3j) with no tag releases in region 5 (implemented as one release and no recaptures).
Rationale: The tag reporting rates for these tag releases are hitting the upper bound of 0.9 and may be restricting the biomass estimates.
Response: This is Run P2. The biomass estimates are substantially higher for this run, but the reporting rates for fisheries 20, 22, 26, and 16-17 are still hitting the 0.9 bound. The tag releases in region 5 are placing an upper bound on the abundance in region 5, but this may be due to limited mixing and partial residence in a restricted region (see Request G). The results of this request were a key part of the basis for the Panel's further investigation of models which exclude these tagging data.

Request P
Request: Construct a variant of the reference model (Run3j) which eliminates some tag releases in region 5 (31, 32) as well as these release groups from region 4 (27, 28, 29) (implemented as one release and no recaptures).
Rationale: As for Request O.
Response: The estimated biomass was even higher for this request than for Request O.

Request Q
Request: As for request P, except also drop release groups from region 3 with reporting rates at the boundary.
Rationale: As for Request O.
Response: This is Run P3. The estimate biomass is substantially larger, and movement patterns change markedly from those for the reference model. This further emphasizes the impact of the tag data on the upper bound for the estimates of biomass.

Request R
Request: Group the weight data into 5kg rather 1kg bins
Rationale: There are a large number of weight-classes and many have few data.
Response: This is Run P4. The results were essentially identical to those for the reference model.

Request S
Request: Fit a Richards model to the size-at-age observations and fix growth in the model
Rationale: The growth curve estimated within the model over-predicts the size-at-at-age observations.
Response: This request could not be completed as the raw data were not available. See Request T for a further request on this topic.

Request T
Request: Fit the reference model in which growth is based on the Richards model.
Rationale: The growth curve estimated within the model over-predicts the size-at-at-age observations.
Response: This is Run P5. There were minor changes to the biomass and recruitment estimates, and a slight reduction in asymptotic length.

Request U
Request: Construct a variant of the reference model that only includes regions 3 and 4
Rationale: Regions 3 and 4 represent the core of the fishery and also contain most of the data. The Panel was interested in understanding whether regions with low catch/data were driving the results.
Response: The biomass estimates from this run were largely comparable with expectations from the reference model which included all regions.

Request V
Request: Repeat Run P2, except drop the Japanese longline length-frequency data.
Rationale: The Panel wished to determine the combined effects of removing these two sources of data (Japanese longline length-frequency data and tag releases in region 5).
Response: The overall biomass was lower than for Run P2. However, there was no obvious improvement in the fit to the weight-frequency data.

Request W
Request: Develop a set of alternative weights for the length- and weight-frequency data.
Rationale: The analysts adjusted the weights for some of the fleets in the reference model, but most of the weights were set to "high".
Response: The analysts developed a weighting scheme (Table E.3) which categorized the remaining data sets into "high", "medium", and "low" reliability, and assigned weight-values of 20, 40 and 100 respectively to those three reliability classes (note that higher weight-values correspond to lesser weight). Length- and weight-frequency data which came from designed sampling programs and had high spatial coverage were assigned the highest weight (20).

Request X
Request: Evaluate the inclusion/exclusion/weighting of the Japanese longline length-frequency data based on a) the reference model, b) Run5 (no early CPUE), and c) the reference model with the discretionary weights (Table E.3) for the composition data.
Rationale: The influence of the length composition data may depend on the other data included in the assessment.
Response: Excluding the Japanese longline length-frequency data from the reference model has very little effect on the biomass estimates, although the estimate of the asymptotic length reduces somewhat. Excluding the Japanese longline length composition data from Run 5 reduces the estimates of biomass for the early years, but those in the last years are not changed substantially. Excluding the Japanese longline length-frequency data from the reference model with the discretionary weights had little impact on the estimates of biomass and recruitment.

Request Y
Request: Evaluate the inclusion/exclusion/weighting of the Japanese longline length-frequency data based on a) the reference model, and b) the reference model with the discretionary weights for the composition data, (Table E.3), except base the analyses on models which exclude the tagging data

which hit bounds and consider scenarios in which selectivity for the Japanese longline fisheries is estimated separately for each region.
Rationale: The influence of the length-frequency data may depend on the other data included in the model.
Response: The largest effect was the choice of which tag data sets to eliminate. Excluding tag groups 31 and 32 had the smallest effect on the biomass estimates whereas dropping tag groups 27, 28, 29, 31, 32 and the recaptures by fishery 26 had the largest influence. The impact of dropping the Japanese longline length-frequency data and estimating region-specific selectivity patterns for the Japanese longline fishery were not consistent (sometimes leading to higher and sometimes lower biomass depending on which tag data sets were ignored), indicating a potential lack of stability in the assessment.

Request Z
Request: Evaluate the exclusion of the early Japanese longline length-frequency data in combination with the early Japanese CPUE data. Show results for removing various sets of tag groups.
Rationale: How data are excluded may interact.
Response: There was again sensitivity to changing the specifications of the assessment, but the changes were again not consistent for each choice of which tag groups to omit.

Request AA
Request: Investigate the influence of the pre-1975 composition data by removing it from the run with no early CPUE data (Run5)
Rationale: The early data may be less reliable and they may influence the estimates of current biomass for recent years.
Response: The biomass is reduced in the early years when the pre-1975 composition data are removed. The post 1975 trend is, however, robust, probably due to the tagging data and because catchability parameters are shared among CPUE time series.

Request AB
Request: Run the reference model (Run3j) and the model with the stock-recruitment curve fit to the later period (Run21) with different weighting factors (related to recruitment deviate cv) for the stock-recruitment relationship.
Rationale: The difference in biomass estimates between Run3j and Run21 was unexpected and may be due to the implicit removal of the penalty on the early recruitment deviates.
Response: The biomass estimates from Run3j were smaller as the penalty on the recruitment deviates was reduced, but the amount to which they were smaller dropped as the weight was further reduced. The biomass estimates from Run21 did not change much when the penalty on the recruitment deviates was reduced. The difference in biomass estimates between Runs3j and 21 was due to the implicit removal of the recruitment deviate penalty when using the shorter period to estimate the stock-recruitment relationship.

Request AC
Request: Use the regional CPUE catchability weights from the aggregated CPUE in the model that uses the operational CPUE (Run3c)
Rationale: The change from Run3b to Run3c included both the use of the operational CPUE data and the regional CPUE catchability weights, and it was unclear which of these led to the difference in the results.
Response: The largest change in the regional CPUE catchability weights was in region 3. The regional CPUE catchability weights had little influence on the total abundance estimates. However, they caused the estimated biomass to move from regions 3 and 4 to regions 1 and 2. The regional CPUE catchability weights sum to one and did not change for region 5 where the tagging data constrains absolute biomass, which probably led to the stability in estimates of total biomass.

Request AD

Request: Run a model that only includes regions 3 and 4, with asymptotic Japanese longline selectivity, no Chinese Taipei composition data, and Chinese Taipei selectivity equal to the Japanese longline selectivity.

Rationale: The results may depend on the Chinese Taipei fishery data, and it is unclear why the fish caught by this fishery are larger than those caught in the Japanese longline fisheries.

Response: The estimated biomass was lower and recruitment increases over time. The estimated asymptotic length is much lower.

Table E.1. Model fisheries as defined, withMCFL weights assigned to and length-, indexselectivity pattern,the index of catchability, *q*

Description (gear, area, fleet)	**Weight Length**	**Weight Weight**	**Selectivity**	***q***	**Region**
LL ALL 1 All Longline	10000	20	1	1	1
LL ALL 2 All, except United States Longline	10000	20	1	1	2
LL HW 2 United States (Hawaii) Longline	20		2	2	2
LL ALL 3 All, except CT-Offshore, CN, FSM, MH, PH, ID, and PNG Longline	10000	20	3	1	3
LL TW-CH 3 CT-Offshore, CN, FSM, MH, PH, and ID	100		4	3	3
LL PG 3 Papua New Guinea Longline	40	40	5	4	3
LL ALL 4 All except CT-Offshore, CN, FSM, MH, and US	10000	20	3	1	4
LL TW-CH 4 CT, CN, FSM, MH, PH, and ID	100	100	3	5	4
LL HW 4 United States (Hawaii) Longline	20	20	6	6	4
LL ALL 5 All except Australia Longline	10000	20	3	1	5
LL AU 5 Australia Longline		40	7	7	5
LL ALL6 All DWFN Longline	10000	20	3	1	6
LL PI 6 Pacific Island Countries/Territories		100	8	8	6
PS ASS 3 All Purse seine, log/FAD sets	20		9	9	3
PS UNS 3 All Purse seine, school sets	20		10	10	3
PS ASS 4 All Purse seine, log/FAD sets	100		9	11	4
PS UNS 4 All Purse seine, school sets	100		10	12	4
PH MISC 3 Philippines Misc. (small fish)	100		11	13	3
PH HL 3 Philippines, Indon. Handline (large fish)	100		12	14	3
PS JP 1 Japan Purse seine	100		13	15	1
PL JP 1 Japan Pole-and-line	100		14	16	1
PL ALL 3 Japan, Solomon's, PNG Pole-and-line	100		15	17	3
LL BMK 3 All, except CT-Offshore, CN, FSM, MH, PH, ID, and PG Longline, Bismarck Sea	10000	20	16	18	3
ID MISC 3 Indonesia Miscellaneous (small fish)	100		11	19	3
HL HW 4 United States (Hawaii) Handline		20	17	20	4
PH-ID PS 3 Philippines, Indonesia- domestic	100		18	21	3

Table E.2. Model fisheries as defined, with the input sample sizes for length and weight frequency (columns 3 and 5) and the effective sample size reference case model.*

Description (gear, area, fleet)	CPUE Input Mean CV	Length Input mean sample size	Length Mean Effective N	Weight frequency Input mean sample size	Weight frequency Mean Effective N
LL ALL 1 All Longline	0.16	19	54	22	165
LL ALL 2 All, except United States Longline	0.22	24	64	16	171
LL HW 2 United States (Hawaii) Longline		20	103		
LL ALL 3 All, except CT-Offshore, CN, FSM, MH, PH, ID, and PNG Longline	0.2	18	65	29	359
LL TW-CH 3 CT-Offshore, CN, FSM, MH, PH, and ID		1	216	1	392
LL PG 3 Papua New Guinea Longline		9	30	33	61
LL ALL 4 All except CT-Offshore, CN, FSM, MH, and US	0.23	28	246	17	395
LL TW-CH 4 CT, CN, FSM, MH, PH, and ID					
LL HW 4 United States (Hawaii) Longline		18	79	50	258
LL ALL 5 All except Australia Longline	0.2	25	69	24	172
LL AU 5 Australia Longline				49	236
LL ALL6 All DWFN Longline	0.35	16	46	7	85
LL PI 6 Pacific Island Countries/Territories		39	184	27	170
PS ASS 3 All Purse seine, log/FAD sets		33	74		
PS UNS 3 All Purse seine, school sets		7	35		
PS ASS 4 All Purse seine, log/FAD sets		27	49		
PS UNS 4 All Purse seine, school sets		5	19		
PH MISC 3 Philippines Misc. (small fish)		21	24		
PH HL 3 Philippines, Indon. Handline (large fish)		5	41		
PS JP 1 Japan Purse seine		24	12		
PL JP 1 Japan Pole-and-line		16	21		
PL ALL 3 Japan, Solomon's, PNG Pole-and-line		5	9		
LL BMK 3 All, except CT-Offshore, CN, FSM, MH, PH, ID, and PG Longline, Bismarck Sea		0	54	17	115
ID MISC 3 Indonesia Miscellaneous (small fish)		12	22		
HL HW 4 United States (Hawaii) Handline				24	84
PH-ID PS 3 Philippines, Indonesia- domestic		1	9		

*mean effective samples computed as: $N_{eff} = \frac{1}{J}\sum_{j=1}^{J}\frac{\sum_{i=1}^{n}\hat{p}_i\left(1-\hat{p}_i\right)}{\sum_{i=1}^{n}\left(p_i-\hat{p}_i\right)^2}$ where $p_i, \hat{p}_i, J$ is are respectively the observed and predicted proportions, and number of years of data.

Pearson residual code

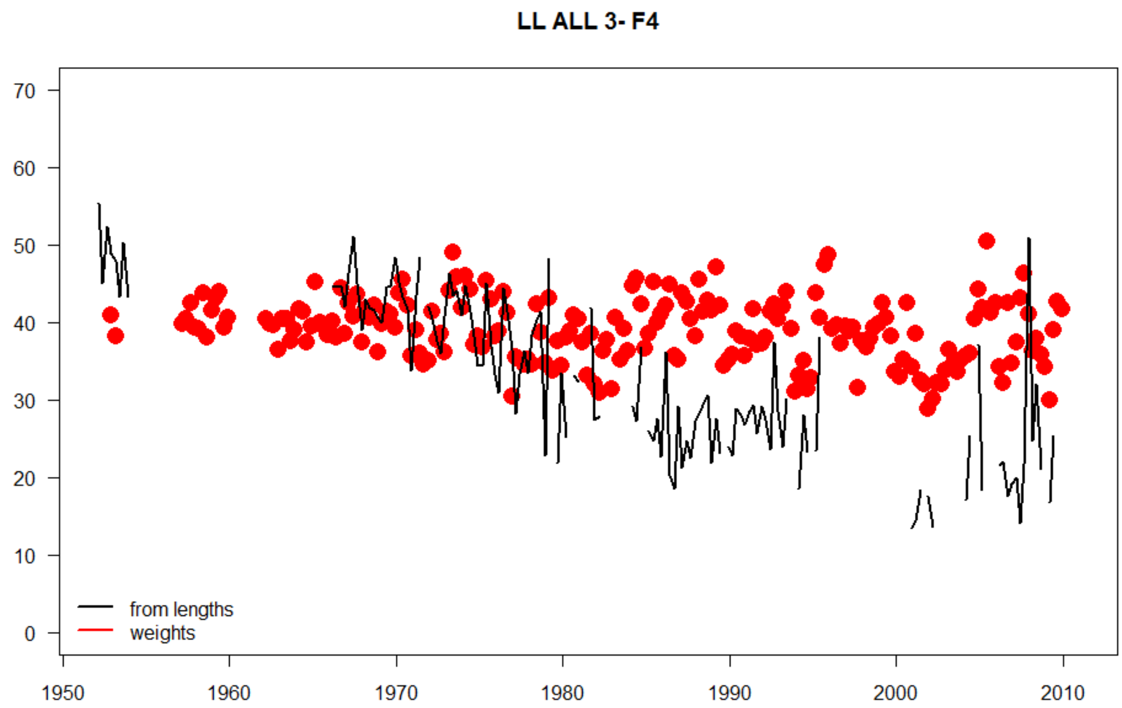
```
get.pears.resid <- function (obslf, predlf, nsamples) {
  pears.resid<-(obslf-predlf)/(sqrt(predlf*(1-predlf)/nsamples))
  return(pears.resid)
}
```

Table E.3. Model fisheries as defined, with the discretionary weights.

Description (gear, area, fleet)	Weight Length	Weight Weight
LL ALL 1 All Longline	10000	20
LL ALL 2 All, except United States Longline	10000	20
LL HW 2 United States (Hawaii) Longline	20	
LL ALL 3 All, except CT-Offshore, CN, FSM, MH, PH, ID, and PNG Longline	10000	20
LL TW-CH 3 CT-Offshore, CN, FSM, MH, PH, and ID	100	
LL PG 3 Papua New Guinea Longline	40	40
LL ALL 4 All except CT-Offshore, CN, FSM, MH, and US	10000	20
LL TW-CH 4 CT, CN, FSM, MH, PH, and ID	100	100
LL HW 4 United States (Hawaii) Longline	20	20
LL ALL 5 All except Australia Longline	10000	20
LL AU 5 Australia Longline		40
LL ALL6 All DWFN Longline	10000	20
LL PI 6 Pacific Island Countries/Territories		100
PS ASS 3 All Purse seine, log/FAD sets	20	
PS UNS 3 All Purse seine, school sets	20	
PS ASS 4 All Purse seine, log/FAD sets	100	
PS UNS 4 All Purse seine, school sets	100	
PH MISC 3 Philippines Misc. (small fish)	100	
PH HL 3 Philippines, Indon. Handline (large fish)	100	
PS JP 1 Japan Purse seine	100	
PL JP 1 Japan Pole-and-line	100	
PL ALL 3 Japan, Solomon's, PNG Pole-and-line	100	
LL BMK 3 All, except CT-Offshore, CN, FSM, MH, PH, ID, and PG Longline, Bismarck Sea	10000	20
ID MISC 3 Indonesia Miscellaneous (small fish)	100	
HL HW 4 United States (Hawaii) Handline		20
PH-ID PS 3 Philippines, Indonesia- domestic	100	

Figure E.1. Observed and mean length-predicted mean weights for one of the Japanese longline fisheries and the Chinese Taipei longline fishery in region 3.

(a) Japanese longline fishery in region 3



(b) Chinese Taipei longline fishery in region 3.