Review of the 2013 Stock Assessment of Pacific Bluefin Tuna[1]

Tom Carruthers

t.carruthers@fisheries.ubc.ca

335 AERL 2202 Main Mall

University of British Columbia

Vancouver

Canada

28th June 2013

[1] PBTWG. 2013. Stock assessment of Pacific Bluefin tuna in 2012. Pacific Bluefin Tuna Working Group of the International Committee for tuna and tuna-like species in the North Pacific Ocean.

1. Executive Summary

The stock assessment has a number of central strengths. Stock Synthesis 3 was used to conduct the assessment which has been subject to extensive review and testing in numerous other stock assessment settings. The Stock Synthesis input files were made available for this review providing a transparent account of data and modelling assumptions. This constitutes best practice and allows reviewers to interact with the assessment to better understand model performance.

There are, however, a number of areas in which the assessment document would benefit from greater detail in order to have confidence in the predictions of the model. I recommend including a wider range of sensitivity analyses, a suitable prior for steepness and using a smaller number of relative abundance indices.

The representative run assumes an extremely high point value for steepness that may have been calculated erroneously. Lower values for this parameter led to model instability and optimization convergence failures. Prescribing a prior for steepness allowed the model to satisfy convergence diagnostics and led to a better overall fit to the data. The same model run estimated a more depleted stock and a substantially higher rate of current exploitation than the representative run.

Maximum likelihood estimates of the representative run were highly sensitive to the assumed level of natural mortality rate (as in other assessments). This may be problematic since the default value for age 2+ fish (0.25 yr^{-1}) appears to be high and implausible given the catch-at-size of certain fleets. Similarly to steepness, the method used to derive natural mortality of age 2+ fish was not subject to peer-review. Both natural mortality rate and steepness estimation methods require detailed description and discussion of assumptions in proportion to their importance for the assessment.

Sensitivity analyses for natural mortality rate and steepness were also limited. The only sensitivity analysis for steepness compared a value of 1.000 to the representative run assumption of 0.999 (0.8 did not converge). The sensitivity analysis for natural mortality rate described sensitivities solely in terms of the impact on estimates of current spawning stock biomass and recruitment in 1990. The same sensitivity analyses were conducted in this review and confirmed these results but also found very high sensitivity in current estimates of exploitation rate. Twenty sensitivity analyses for data weightings, relative abundance indices and vulnerability schedules were conducted. However, these covered a relatively narrow range of possible model and data configurations. For example, as part of this review an assessment based solely on the Japanese longline relative abundance indices produced reference point estimates that differed from the representative run to a greater extent than any of the twenty sensitivity runs.

The assessment makes use of several CPUE derived indices that are likely to provide an unreliable inference of relative abundance changes such as nominal purse-seine CPUE. By incorporating multiple relative abundance indices that provide contradictory inferences about abundance trends, the assessment appears to down-weight these data. The product is that the assessment depends to a greater extent on catch composition data and total catches to inform stock depletion. This is concerning since the model fails to approximate several sources of size composition data and generates patterns in residuals that indicate model mis-specification. The majority of the parameters for vulnerability schedules were user-specified and did not appear to allow the model to fit the composition data for several fleets.

Posterior parameter cross-correlation plots for a Bayesian analysis of the representative run indicate parameter confounding and a poorly defined objective surface. Changing the order of phases in which parameters were estimated led to different estimates of reference points whilst satisfying convergence

diagnostics of SS3. It is therefore difficult to confirm convergence to a global minimum in the representative run. The protocol by which estimation phases were assigned to estimated parameters was not described.

It is not clear why it was necessary to explicitly model seasonal population dynamics. The rationale for modelling 14 different fishing fleets is also not clear. This level of fleet disaggregation appears to be overly complex since it explicitly models the vulnerability schedules of several minor fleets, one of which constitutes just 0.03% of the historical catches of Pacific bluefin tuna.

In general, the assessment results are difficult to interpret in terms of the status of the stock relative to productive levels (*e.g.* B_{MSY}) and it is hard to gain an intuition for the likely stock trajectory at current exploitation levels (*e.g.* relative to F_{MSY}). Stock projections did not appear to correspond with other stock assessment estimates. For example fishing mortality rate in the period 2002-2004 was estimated to be 2.5 times $F_{0.1}$ and 1.75 times F_{MAX} and yet was predicted to lead to increases in future spawning stock biomass.

More clearly defined management objectives for the Pacific bluefin stock would enable future assessment to be presented in a meaningful framework and support the development of quantitative tools for decision making (*e.g.* MSE).

2. Background to the review

Pacific bluefin tuna (*Thunnus orientalis*) is one of three species of bluefin tuna that also includes two stocks in the Atlantic Ocean (*Thunnus thynnus*) and another inhabiting the Southern hemisphere of all Oceans (*Thunnus maccoyii*). In both their methodology, results and uncertainties the assessments of these stocks provide a context for the work under review here. According to the most recent assessments, global bluefin tuna stocks are at low levels but have a mixed outlook based on current exploitation levels (ICCAT 2012, CSSBT 2011). Similarly to Pacific bluefin tuna, Southern Bluefin Tuna (SBT) are considered to be heavily depleted at around 3-7% of unfished levels but unlike the Pacific stock, SBT is thought to be currently underfished (CCSBT 2011). There are large differences in both the complexity of assessments for these stocks and way in which uncertainty is accounted for in management decision making. Relatively simple VPA analyses are applied to the two stocks of the Atlantic while SBT is assessed by statistical catch-at age supported by Management Strategy Evaluation (MSE).

The model structure of the current assessment is similar to previous assessments. The most recent of which (PBTWG 2010) also used Stock Synthesis 3 to conduct statistical catch-at-age analysis in which a single mixed Pacific stock was assumed and fishing was approximated by a similar disaggregation of fleets. This followed a similar approach applied in Stock Synthesis 2 in 2008 (PBTWG 2008). Among the most important changes to the 2010 assessment was the shift towards higher natural mortality rate of adult fish (age 2+) from $0.12yr^{-1}$ to $0.25yr^{-1}$ in order to address apparent model misspecification. The 2010 assessment predicted recent increases in fishing mortality rate and continual decline in spawning stock biomass at ‘current’ fishing mortality levels.

Of the stocks described above, the fishery dynamics are arguably the most complicated for Pacific bluefin. The stock has been subject to exploitation by many fleets (flag and gear combinations) of varying size selectivity. Several of these fleets have operated intermittently. Some of these fleets are regional and interact with a different subdivision of the Pacific Ocean population due to spatial heterogeneity in the stock and/or seasonal changes in population distribution. The complexity of these interactions between fishing and the population poses difficulties for stock assessment using statistical catch-at-age analysis that aims to approximate the vulnerability-at-age of fishing fleets. The assessment is complicated further by the presence of important fisheries prior to 1952, a period in which among the highest annual catches are thought to have occurred. Since detailed data on size composition, relative abundance and catches are only available after a period of substantial stock reduction (*e.g.* spawning stock biomass in 1952 at around 10% of unfished levels) these data may be considerably less informative about stock size and productivity.

One of the most important and challenging aspects of Pacific Bluefin tuna assessment is accounting for spatial phenomena. On the one hand there is evidence of a single genetically mixed stock throughout the Pacific (Tseng and Smith 2011) and tagging studies have confirmed trans-oceanic movements (*e.g.* Boustany *et al*. 2010). On the other hand the same tagging studies provide evidence for considerable viscosity in the stock that could lead to regional depletions which cannot be approximated by a model that assumes that the stock is fully spatially mixed. The approach used here and in other tuna assessments is to implicitly account for spatial effects using multiple fleets of different vulnerability schedules. Such a model may be successful at accounting for general differences in regional size composition and density but may not account well for changes in the fleet relative to stock density over time (for example spatial targeting leading to fishing in higher density areas) and cannot account for regional depletions.

Assessment model parsimony is a central concern when considering how to account for seasonality, spatial heterogeneity and changes in fishing behavior within and among fleets. The assessment must include sufficient detail (*e.g*. disaggregation of fleets, temporal changes in vulnerability) to reliably inform stock size, productivity and current exploitation characteristics for use in decision making. However, it may be difficult to diagnose a model that is overparameterized in which reference points are not informed reliably by the data.

Pacific Bluefin tuna are estimated to be heavily depleted at around 3-4% of unfished stock size. This level of stock depletion can make assessment results particularly sensitive to inputs for key population parameters such as steepness and natural mortality rate. Since these inputs are generally difficult to determine empirically it can prove difficult to decide on a default model structure (‘base case’ assessment or ‘representative run’). Expressing uncertainty in these inputs in terms of model predictions is therefore critical.

3. Description of the Individual Reviewers Role in the Review Activities

A detailed description of the reviewer’s role can be found in the Statement of Work (Appendix 2). The supporting documents and main assessment report were received on the 14th May 2013.

The control, starter and data files of the SS3 assessment were made available providing a detailed account of the model assumptions and data. This degree of transparency represents best practice. By allowing the reviewer to interactively interrogate the representative run, guesswork is greatly reduced: the reviewer can gain an intuitive understanding of model behaviour and performance and answer questions as they arise. Additionally, graphics and tables can be produced to explain points that are made by the reviewers. To this end a number of sensitivity runs were undertaken as part of this review based on alternative scenarios for natural mortality rate, steepness, equilibrium fishing conditions, the relative abundance data and the order in which parameters were admitted into the optimization. The results of these sensitivity runs are included in Table R1 and Figure R1. It should be noted that these sensitivity runs were undertaken using the latest version of Stock Synthesis (V3.24f) Windows 64bit (3/10/2012) which provides very slightly different estimates of quantities to the version used in the assessment (V3.23b Windows 64bit). For example, the representative run of the assessment (V3.23b) converges on a total objective function value of 2200.42 (from Figure 5-3) and an estimate of current biomass of 22,606mt (Table 5-5) compared with value of 2200.41 and 22,619mt for the representative run using the latest version of SS3 (V3.24f).

Table R1. The sensitivity of maximum likelihood estimates of reference points to different assumptions about steepness, natural mortality rate, equilibrium catch conditions, relative abundance indices and the allocation of estimation phases. Four natural mortality rate assumptions were tested: mortality at age at 80%, 90%, 110% and 120% of the representative run (M_08, M_09, M_11, M_12, respectively). Four fixed steepness runs were tested h=0.99, h=0.98, h=0.97 and h=0.96 (h99, h98, h97, h96, respectively). Steepness values below 0.96 did not converge. An additional run was undertaken with a bounded log-normal prior for steepness (hprior). Two equilibrium fishing mortality rate scenarios were considered that added the catch of the Japanese purse seine Pacific Ocean (Fleet 4, Eqm1 and Eqm2) and the Japanese pole and line fleets (Fleet 6, Eqm2) to the standard assumption of the representative run, in which Japanese longline (Fleet 1) and Japanese Troll (Fleet 5) vulnerabilities determine the equilibrium fishing mortality rate. Three sensitivity runs are included that removed the purse seine relative abundance series (noPSi), only use the Japanese longline series' (JPLLi) in addition to a run that stitches together the three Japanese longline indices (JPstiti). A further three sensitivity runs examined the effect of estimation phase on results in which all selectivity parameters were estimated in phase 1, 2 and 3 of the optimization (SelPh1, SelPh2, SelPh3, respectively).

Run	Depletion	B_{2010}/B_{MSY}	F_{2010}/F_{MSY}	SSB_0	SSB_{2010}	Depletion $_{1952}$	Objective F
Representative	3.57%	18%	155%	633468	22619	8.95%	2200.41
M_12	6.36%	39%	104%	416781	26507	13.94%	2209.48
M_11	4.85%	27%	127%	508561	24657	11.26%	2204.39
M_09	2.53%	12%	188%	807343	20416	7.27%	2197.46
M_08	1.70%	8%	231%	1054820	17969	6.58%	2195.46
h99	3.43%	17%	159%	655428	22457	8.72%	2200.08
h98	3.28%	16%	164%	682100	22392	7.86%	2200.08
h97	3.13%	15%	169%	712097	22308	8.22%	2199.53
h96	3.00%	14%	173%	744700	22345	7.96%	2199.34
hprior	2.81%	12%	181%	809455	22709	7.87%	2199.21
Eqm1	3.69%	19%	154%	634876	23431	14.94%	2202.2
Eqm2	3.81%	20%	152%	637793	24284	21.37%	2205.59
JPLLi	3.47%	18%	203%	623031	21649	7.99%	2227.14
noPSi	3.57%	18%	155%	633468	22619	8.95%	2200.41
JPstiti	2.78%	14%	238%	616706	17141	17.66%	2235.42
SelPh1	3.45%	17%	160%	629419	21701	9.82%	2200.86
SelPh2	3.57%	18%	155%	633468	22619	8.95%	2200.41
SelPh3	3.58%	18%	154%	633683	22696	8.95%	2200.91

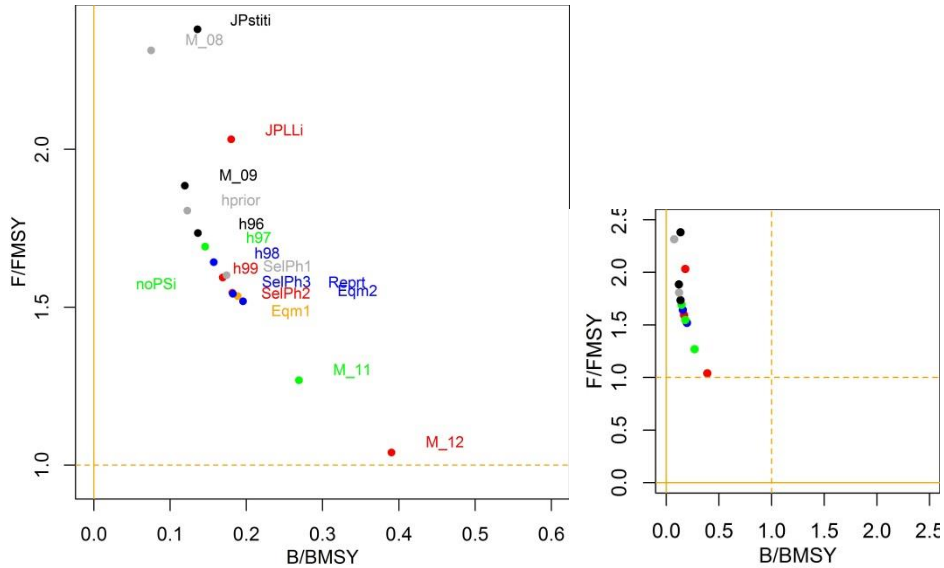

Figure R1. Sensitivity of maximum likelihood estimates of current (2010) exploitation levels (F/F_{MSY}) and stock status (B/B_{MSY}) based on natural mortality rate and steepness inputs (see Table R1 caption for a description of these runs).

4. Summary of findings in regard to TORs (weaknesses and strengths)

4.1. Review the assessment methods to provide recommendations on how to improve its application, and/or recommend other methods that would also be appropriate for the species, fisheries, and available data.

A specific discussion of these issues is included in Section 4.2 below. It is difficult to have confidence over the assessment results given (1) the use of a steepness value that may not have been derived correctly, (2) failure of the model to converge when using alternative credible values for steepness, (3) very high sensitivity in model outputs when using a smaller range of relative abundance indices that may be more representative of overall changes in stock trend, (4) failure of the model to approximate the catch composition of numerous fleets, (5) difficulty in confirming convergence to a global minimum and (6) projections that appear inconsistent with other model estimates. For the benefit of simplicity recommendations are summarised in a separate Section (Section 6).

4.2. Evaluate the assessment model configuration, assumptions, and input parameters (fishery, life history, and spawner recruit relationships) to provide recommendations on how to improve: the use of data, specification of fixed input parameters, and specification of model configuration.

4.2.1. Model configuration and assumptions

Model misspecification and the inclusion of minor fishing fleets

The representative run attempts to estimate the vulnerability schedules for 14 separate fleets. However 80% of the historical catches were taken from just five fleets (Table R2). Four different Japanese set net fisheries are described with different seasonal contributions that each amount to about 2.5% of the historical catches. Smaller still is the Eastern Pacific sports fishery which caught just 0.03% of the historical catch of bluefin. The justification for including these minor fleets is not clear.

Table R2. Historical catches of the fleets of the representative run.

Fleet number	Fleet	% Historical catches	% Cumulative	Alternative fleet assignment
4	**Tuna PS Pacific Ocean**	30.3%	30.3%	1
12	**Eastern Pacific commerc**	24.0%	54.4%	2
2	**Small pelagic PS**	9.3%	63.7%	3
5	**Japanese coastal troll**	9.2%	72.9%	4
1	**Japanese longline**	6.7%	79.6%	5
6	**Japanese pole and line**	5.6%	85.1%	6
9	**Japanese set net**	2.9%	88.0%	7
7	**Japanese set net**	2.3%	90.3%	7
10	**Japanese set net**	2.1%	92.4%	7
3	**Tuna PS Sea of Japan**	2.1%	94.5%	8
11	**Taiwanese longline**	2.1%	96.6%	8
8	**Japanese set net**	1.8%	98.4%	7
14	**other**	1.5%	100.0%	8
13	**Eastern Pacific sports**	0.0%	100.0%	8

The principal concern is that the minor fisheries may contribute substantially to the likelihood component associated with the composition data and possibly compromise the fit to the data of the more important fleets that may also have more reliable composition data. This may be important since there are indications that the composition data may be particularly influential in this assessment. Alternative sensitivity runs indicate that the relative abundance indices do not strongly determine the estimate of current depletion (Section 'Initial conditions and equilibrium exploitation' below, runs Eqm1 and Eqm2, Table R1), implying that the inference regarding depletion is instead coming from the size composition data. This relies heavily on the quality of these data. It is concerning therefore that the assessment appears to generate problematic patterns in the residuals of the size composition data[2].

There are several examples of fleets where selectivity schedules fail to fit a large band of size observations above a certain length. For example, Fleets 7, 9 and 14 (Japanese set net and other gears) that show consistent patterns of positive residuals above size thresholds of 50cm, 100cm and 200cm, respectively. At first glance it is not clear why the optimizer should not identify parameters for the descending limb that would fit these data. However, in Section 4.3.2 of the assessment document it is stated that:

> *In this assessment, selectivity patterns were estimated for all fisheries with length composition data except for Fleet 14, which was a composite of multiple different gears, and Fleet 6, which was poorly sampled relative to a similar fishery (Fleet 5).* (Section 4.3.2, page 29)

There are additional indications that some other selectivity parameters are also user-specified:

> *For most fisheries, the initial and final parameters of the selectivity patterns were assigned values of -999 or fixed to a small value (-15), which caused SS to ignore the first and last size bins and allowed SS to decay the small and large fish selectivity according to parameters of*

[2] The legend of Figure 5-5 of the assessment document is ambiguous: "Dark blue circles indicate negative residuals (observation value < expected value), while white circles indicate positive residuals (expected value > observation value)". When commenting on residual error, I assumed the traditional definition of residuals (observed - expected) which is also consistent with Figures 5-22 and 5-23 of the assessment document.

> *ascending width and descending width, respectively. For some fisheries, the parameter specifying the width of the plateau was often estimated to be very small (-9) and often hit assigned bounds. For these fisheries, the width of the plateau was set to -9. Other parameters describing domed-shape selectivity were estimated by the model, i.e., beginning size for the plateau, ascending width, and descending width.* (Section 4.3.3., paragraph 3, page 30)

The control file for the assessment provides further insight. Of the 72 parameters that describe the selectivity distributions, 34 are estimated and 38 are user-defined (Table R3). While the report provides some rationale for the values of these fixed parameters there is evidence that these are not appropriate and could be adjusted to remove problematic patterns in the residual errors.

These points raise a number of questions that are not well addressed in the assessment document. For example, how was the decision made to fix some parameters and allow others to be estimated by Autodiff? If these parameters are not well informed by the data, what does this indicate about model complexity? By what protocol are the values for user-specified selectivity parameters assigned given the relatively poor fit to the observed data? Perhaps most importantly, why should the assessment make use of size composition data for minor fleets that in any case are poorly predicted by the model? This point is made regarding Fleet 14 but I would argue it also applies to several other fleets that are modelled individually in this assessment:

> *Given the relatively small catches from this fleet and the difficulties in modeling the selectivity of this fleet, the selectivity of Fleet 14 was fixed with parameters estimated by a preliminary run with lambda=0.1.* (Section 4.3.4. Second paragraph, page 30)

Additionally there are fleets for which blocked selectivities are estimated for a very small number of recent years. For example the Japanese purse seine fishery (Fleet 3) includes a blocked selectivity for the most recent four years (2007-2010). There may be observations to support the hypothesis that the fishery operated differently before and after 2007 (*e.g*. Fukuda et al. 2012) but this selectivity may not be reliably estimated over such a short time period and could have a relatively strong impact on stock projections. It would be instructive to see the sensitivity of the model to the removal of this selectivity block.

In my view there is too much assessment machinery associated with approximating the selectivities of small fleets. A simpler model may provide a better basis for reliable decision making. A much more detailed rationale for this relatively complex fishery structure is required. Simulation evaluation could be used to identify how best to aggregate the various fleets.

Table R3. The estimated (highlighted in blue) and user-specified selectivity parameters of the representative run. Parameters with negative phase attributes are not estimated and instead fixed to the initialization value (INIT).

Selectivity parameter	LO	HI	INIT	PRIOR	PR_type	SD	PHASE
SizeSel_1P_1_F1JLL	21.2	284.1	205.324	44.8675	0	999	2
SizeSel_1P_2_F1JLL	21.2	284.1	205.324	44.8675	0	999	-2
SizeSel_1P_3_F1JLL	-1	9	7.18608	2.33625	0	999	3
SizeSel_1P_4_F1JLL	-1	9	5.44148	7.20306	0	999	4
SizeSel_1P_5_F1JLL	-999	-999	-999	-5	0	999	-4
SizeSel_1P_6_F1JLL	-5	9	-2.85728	-5	0	999	4
SizeSel_2P_1_F2SPelPS	21.2	284.1	49.1704	44.8675	0	999	2
SizeSel_2P_2_F2SPelPS	-9	4	-9	-6	0	999	-2
SizeSel_2P_3_F2SPelPS	-1	9	3.53812	2.33625	0	999	3
SizeSel_2P_4_F2SPelPS	-1	9	6.34505	7.20306	0	999	4
SizeSel_2P_5_F2SPelPS	-999	-999	-999	-5	0	999	-4
SizeSel_2P_6_F2SPelPS	-999	-999	-999	-5	0	999	-4
SizeSel_3P_1_F3TunaPSJS	21.2	284.1	139.591	46.6534	0	999	2
SizeSel_3P_2_F3TunaPSJS	-9	10	-8.52129	-5	0	999	2
SizeSel_3P_3_F3TunaPSJS	-1	10	5.71211	5.35469	0	999	3
SizeSel_3P_4_F3TunaPSJS	-1	10	8.52121	6.75769	0	999	2
SizeSel_3P_5_F3TunaPSJS	-999	-999	-999	-5	0	999	-4
SizeSel_3P_6_F3TunaPSJS	-999	-999	-999	-5	0	999	-4
SizeSel_4P_1_F4TunaPSPO	21.2	200.1	107.262	46.6534	0	999	2
SizeSel_4P_2_F4TunaPSPO	-9	9	-8.32833	-6	0	999	2
SizeSel_4P_3_F4TunaPSPO	-9	12	6.85151	5.35469	0	999	3
SizeSel_4P_4_F4TunaPSPO	-9	12	9.02474	6.75769	0	999	2
SizeSel_4P_5_F4TunaPSPO	-999	-999	-999	-5	0	999	-4
SizeSel_4P_6_F4TunaPSPO	-999	-999	-999	-5	0	999	-4
SizeSel_5P_1_F5JpnTroll	21.2	284.1	52.9001	46.6534	0	999	2
SizeSel_5P_2_F5JpnTroll	-9	4	-9	-9	0	999	-2
SizeSel_5P_3_F5JpnTroll	-1	9	6.03766	5.35469	0	999	3
SizeSel_5P_4_F5JpnTroll	-1	9	4.64663	6.75769	0	999	2
SizeSel_5P_5_F5JpnTroll	-999	-999	-999	-5	0	999	-4
SizeSel_5P_6_F5JpnTroll	-999	-999	-999	-5	0	999	-4
SizeSel_6P_1_F6JpnPL	1	14	1	1	0	25	-99
SizeSel_6P_2_F6JpnPL	-5	0	-1	-1	0	25	-99
SizeSel_7P_1_F7JpnSetNetNOJWeight	21.2	284.1	82.041	46.6534	0	999	2
SizeSel_7P_2_F7JpnSetNetNOJWeight	-9	4	-9	-6	0	999	-2
SizeSel_7P_3_F7JpnSetNetNOJWeight	-1	9	6.11136	-5	0	999	3
SizeSel_7P_4_F7JpnSetNetNOJWeight	-1	9	6.94831	6.75769	0	999	2
SizeSel_7P_5_F7JpnSetNetNOJWeight	-15	9	-15	-5	0	999	-4
SizeSel_7P_6_F7JpnSetNetNOJWeight	-15	9	-15	-5	0	999	-4
SizeSel_8P_1_F8JpnSetNetNOJLength	21.2	284.1	81.8371	46.6534	0	999	2
SizeSel_8P_2_F8JpnSetNetNOJLength	-9	9	-9	-3	0	999	-4
SizeSel_8P_3_F8JpnSetNetNOJLength	-1	15	14.1163	5.35469	0	999	3
SizeSel_8P_4_F8JpnSetNetNOJLength	-1	12	7.61891	6.75769	0	999	2
SizeSel_8P_5_F8JpnSetNetNOJLength	-15	9	-15	-5	0	999	-4
SizeSel_8P_6_F8JpnSetNetNOJLength	-15	9	-15	-5	0	999	-4
SizeSel_9P_1_F9JpnSetNetOAJLength_Q123	21.2	284.1	77.2085	46.6534	0	999	2
SizeSel_9P_2_F9JpnSetNetOAJLength_Q123	-9	4	-9	-6	0	999	-2
SizeSel_9P_3_F9JpnSetNetOAJLength_Q123	-1	9	6.54703	5.35469	0	999	3
SizeSel_9P_4_F9JpnSetNetOAJLength_Q123	-1	9	6.32998	6.75769	0	999	2
SizeSel_9P_5_F9JpnSetNetOAJLength_Q123	-15	9	-15	-5	0	999	-4
SizeSel_9P_6_F9JpnSetNetOAJLength_Q123	-15	9	-15	-5	0	999	-4
SizeSel_10P_1_F10JpnSetNetOAJLength_Q4	21.2	284.1	65.0592	46.6534	0	999	2
SizeSel_10P_2_F10JpnSetNetOAJLength_Q4	-6	9	-6	-6	0	999	-2
SizeSel_10P_3_F10JpnSetNetOAJLength_Q4	-1	9	4.52398	5.35469	0	999	3
SizeSel_10P_4_F10JpnSetNetOAJLength_Q4	-1	9	8.9999	6.75769	0	999	-2
SizeSel_10P_5_F10JpnSetNetOAJLength_Q4	-15	9	-15	-5	0	999	-4
SizeSel_10P_6_F10JpnSetNetOAJLength_Q4	-15	9	-15	-5	0	999	-4
SizeSel_11P_1_F11TWLL	60	230	212.56	206.5	0	999	2
SizeSel_11P_2_F11TWLL	0.1	50	17.737	40	0	999	3
SizeSel_12P_1_F12EPOPS	21.2	150.1	68.3123	73.8791	0	999	2
SizeSel_12P_2_F12EPOPS	-9	4	-9	-3.81522	0	999	-3
SizeSel_12P_3_F12EPOPS	-1	9	4.15302	4.689	0	999	4
SizeSel_12P_4_F12EPOPS	-1	9	6.92287	6.07867	0	999	5
SizeSel_12P_5_F12EPOPS	-1014	-1014	-1014	-5	0	999	-4
SizeSel_12P_6_F12EPOPS	-15	9	-15	-5	0	999	-4
SizeSel_13P_1_F13EPOSP	1	14	1	1	0	25	-99
SizeSel_13P_2_F13EPOSP	-5	0	-1	-1	0	25	-99
SizeSel_14P_1_F14others	21.2	284.1	59.4794	46.6534	0	999	-2
SizeSel_14P_2_F14others	-6	9	-1	-2.5	0	999	-4
SizeSel_14P_3_F14others	-1	10	4.45828	5.35469	0	999	-3
SizeSel_14P_4_F14others	-1	10	4.73724	6.75769	0	999	-2
SizeSel_14P_5_F14others	-1	-1	-1	-5	0	999	-4
SizeSel_14P_6_F14others	-9	9	8.03418	-5	0	999	-4

Initial conditions and equilibrium exploitation

The model uses the selectivity schedules of the Japanese longline (Fleet 1) and Japanese troll (Fleet 5) fleets to initialize the model. This decision is justified by:

> *..these two fleet were chose to estimate initial Fs because they represented fleets that take large and small fish.* (Section 4.3.1., page 29)

The representative run estimates that in 1952 the stock was in an overfished state and subject to overfishing (Figure R2). This may be true but it is a difficult starting point from which to conduct statistical catch-at-age modelling. The most serious problem is that the detailed data that are available provide a much weaker inference regarding reference points for current exploitation rate and stock level. It also raises the question: how sensitive are the predictions of the representative run to changes in the initialization assumptions (*e.g.* the fleets that were chosen for the equilibrium conditions)?

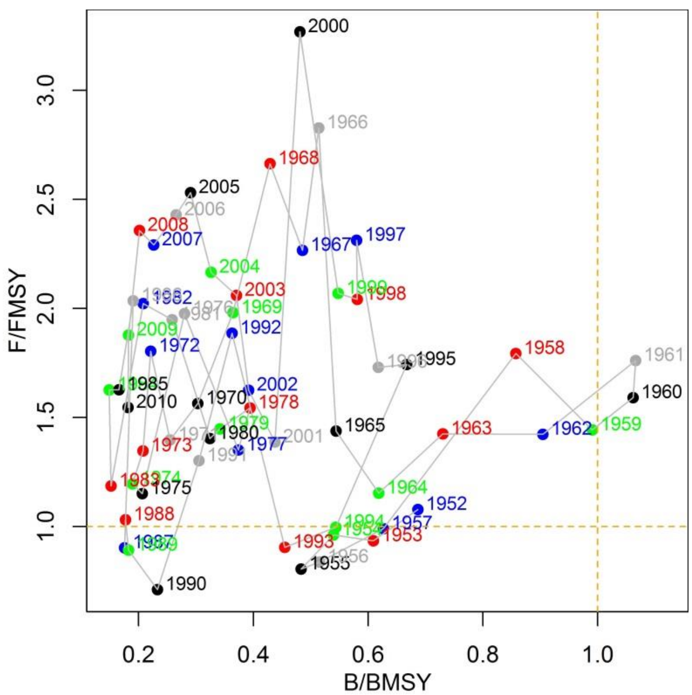

Figure R2. Maximum likelihood estimates of historical exploitation levels (F/F_{MSY}) and stock status (B/B_{MSY}) (B refers to spawning stock biomass) from the representative run.

To investigate initial conditions, two equilibrium fishing mortality rate scenarios were considered that added fishing by the Japanese purse seine Pacific Ocean fleet (Fleet 4, Eqm1 and Eqm2) and the Japanese pole and line fleet (Fleet 6, Eqm2) (Tables R4 and R5). While the effect on initial depletion estimates (depletion in 1952, Table R1) is relatively large, the effect on current reference points is negligible, including current depletion. This indicates that the relative abundance indices may not be driving estimates of current stock depletion relative to 1952. An explanation for this may lie in either (1) the use of non-contiguous survey indices that sever inference of long-term depletion and/or (2) the use of multiple survey abundance indices with contradictory trends.

Table R4. The initial equilibrium catch assumptions of the model (relative to catches of those fleets) and alternative sensitivity runs that use other important fisheries (according to 1952 catches).

Catch	Japanese longline	Small pel PS	Tuna PS sea of japan	tuna PS pacific ocean	Japanese coastal troll	JP Pole and Line	Japanese set net	Japanese set net	Japanese set net	Japanese set net	Taiwanese longline	Easter pacific	Easter pacific sports	other	Total	% of 1952 total catch
Catch in 1952 (mt)	3247	0	0	6926	843	2920	424	498	825	555	0	1975	0	172	**18384**	
% total in 1952	18%	0%	0%	38%	5%	16%	2%	3%	4%	3%	0%	11%	0%	1%		**100%**
Total catches equilibrium assumption for initial F (t)																
Representative run	3247				843										**4090**	**22%**
Eqm1 sense run	3247			6926	843										**11016**	**60%**
Eqm2 sense run	3247			6926	843	2920									**13936**	**76%**
Equilibrium catches assumed																
Representative run	1271				2442										**3713**	**20%**
Eqm1 sense run	1094			2334	284										**3713**	**20%**
Eqm2 sense run	865			1845	225	778									**3713**	**20%**

Table R5. The alternative sensitivity runs for equilibrium conditions. Eqm1 and Eqm2 use initial values and bounds for equilibrium fishing mortality rate that are consistent with the catch apportionment of Table R4 and the total mortality rates for initial values and bounds in the representative run.

Fleet	LO	HI	INIT	PRIOR	PR_type	SD
Representative run/ Eqm1						
InitF_1F1JLL	0	4.9	0.472323	0.384903	0	10000
InitF_5F5JpnTroll	0	10.9	10.2491	1.48535	0	10000
Eqm1						
InitF_1F1JLL	0	4.66	3.16	0.5	0	10000
InitF_4F4TunaPSPO	0	9.93	1.21	0.5	0	10000
InitF_6F6JpnPL	0	6.74	0.82	0.5	0	10000
Eqm2						
InitF_1F1JLL	0	3.68	2.5	0.5	0	10000
InitF_4F4TunaPSPO	0	7.85	5.33	0.5	0	10000
InitF_5F5JpnTroll	0	0.96	0.65	0.5	0	10000
InitF_6F6JpnPL	0	3.31	2.25	0.5	0	10000

A seasonal population dynamics model

From the assessment document it is hard to completely understand the rationale for a seasonally explicit population dynamics model. In several instances it is argued that seasonality is important in determining size composition of catches (it would be useful to include figures or tables to support this assertion). Let us assume that this is valid. The assessment approach here proceeds to disaggregate fleets by season, dividing up the size composition data and catch data to account for differences in availability of individuals of different size. This provides the basis for implicitly accounting for seasonality which is a reasonable approach. It does not follow that it is also necessary to explicitly model abundance by season: the same removals, catch composition and vulnerability schedules of fleets (now named by season) could be applied to a model aggregated by year. A more complete justification for this additional model structure is necessary. It would be instructive to see the sensitivity in estimates of reference points to an otherwise identical assessment model that assumes annual population dynamics.

Convergence and convergence diagnostics

It is concerning that the model failed to converge when fixed steepness values of less than 0.96 were prescribed (which are credible given Mangel *et al*. 2010 and Ikawa 2012), despite improvement in the global objective function with declining steepness (Table R1). There also appears to be cause for concern when examining parameter cross-correlation plots (Figure R3) from an MCMC run of the representative model which indicates potentially serious parameter confounding.

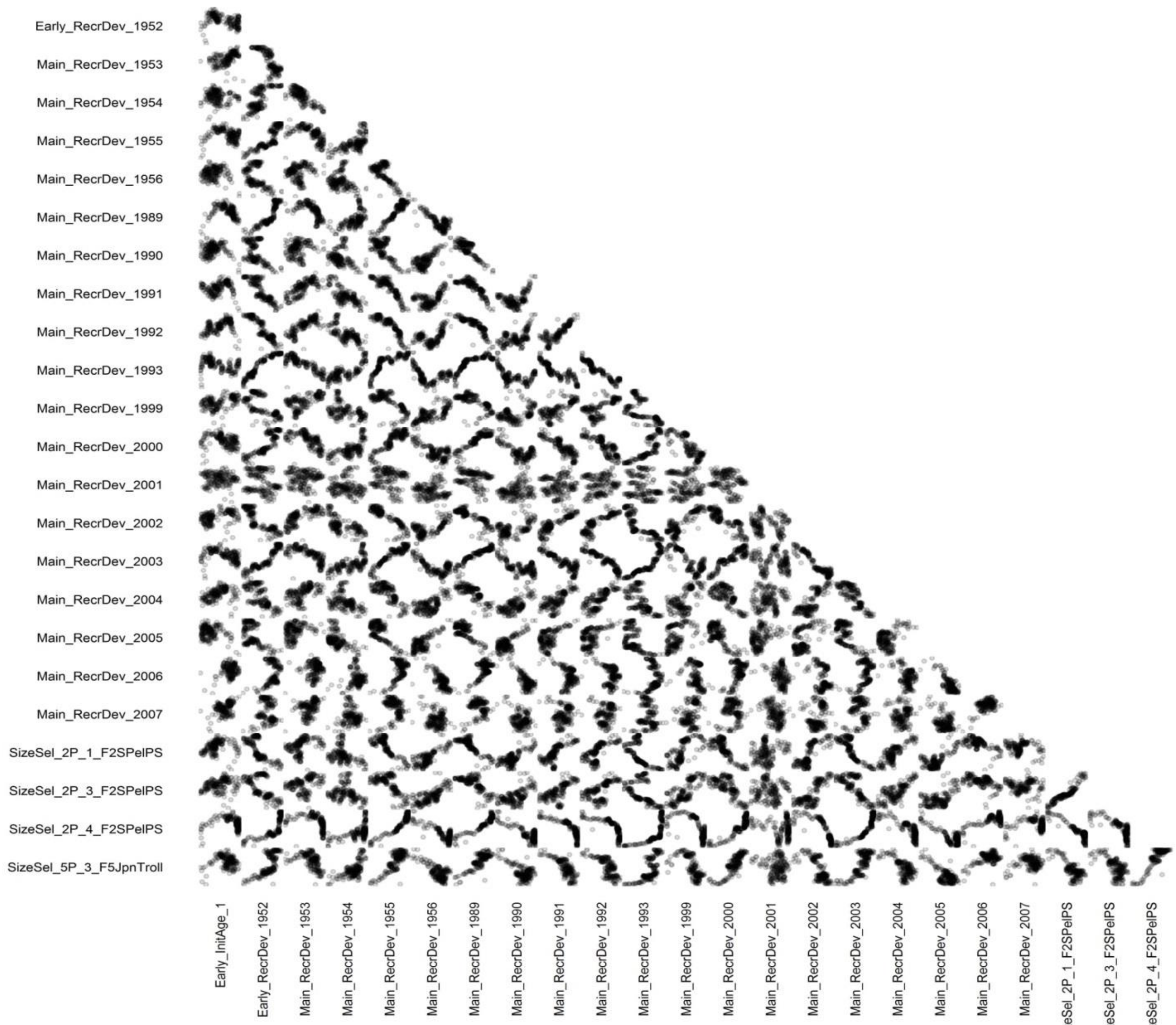

Figure R3. Posterior correlation of a sample of 24 parameters from a Bayesian run of the representative assessment. Parameters were sampled every 1000 iterations over a total of 200,000 MCMC iterations.

It is likely that the objective surface of the representative run (as illustrated in Figure R3) is irregular and has many local minima. The reason for the apparent convergence of the methods based on the jittered runs may in fact be due to a multiple-phased parameter estimation approach (where the optimizer includes a greater number of parameters in the optimization in successive phases) rather than a well-determined estimation problem. A particular phase structure can lead to initial restriction in a parameter space close to a local minima that may be reliably located under successive jittered starting points.

In this review three sensitivity runs were conducted to examine whether the phased estimation was responsible for finding a local minima and creating the appearance of a well-defined optimization problem. In these optimization runs, all of the selectivity parameters were estimated in phases 1, 2 and 3, respectively (SelPh1, SelPh2, SelPh3). These sensitivity runs indicate that the representative run may be locating a parameter vector at a relatively low (good fit) global objective value since selectivity parameters estimated in phases 1 and 3 received larger final objective scores (Table R1) and estimation in phase 2 obtained the same objective score as the representative run. However, these runs confirm that the estimation problem may be poorly defined and the results are dependent on the user-specification of phases. It may be the case that by manual iteration, estimation phases were chosen that reliably corresponded to low total objective function value. However, unlike the Autodiff routines of SS3 this is not a reproducible method and to a certain extent it undermines the jitter analysis as a test of convergence.

It may also be the case that a better overall fit to the data occurs at a different parameter vector that is not easily located by manual alteration of estimation phases. In these instances an MCMC run offers a way of characterising a more complete range of credible parameter vectors whilst simultaneously expressing these in terms of management reference points. In any case, a description of the manual method used to select the phases for estimation is necessary and where possible this process should be included in the sensitivity analysis as it may affect estimates of reference points (current F/F_{MSY} was 160%, 155% and 154% in the SelPh1, SelPh2 and SelPh3 sensitivity runs respectively).

The assessment report provides a brief summary of the results of the jitter analysis of convergence.

> *The jitter runs showed that the model likely converged to a global minimum, with no evidence of further improvements to the total likelihood or substantial trends in the scaling parameter (R0) (Figure 5-3).* (Section 5.1.1. page 36)

Under jittered runs the model did indeed locate an identical parameter vector corresponding to the same objective function score. However as stated above, this may not indicate "that the model likely converged to a global minimum". It should be noted that unfished recruitment (*R0*) and the objective function value are not necessarily suitable metrics for evaluating convergence from different initial values. Even in heavily over-parameterised stock assessment models, unfished stock size (*e.g. R0*) can be well characterized compared with other parameters. In the same instances there can be a very flat, poorly defined objective surface where many different parameter vectors can predict the data similarly well (maintaining similar total likelihoods). In future tests of this kind it would be more insightful to focus on metrics of management interest, for example the extent to which MLEs of F_{2010}/F_{MSY} and SSB_{2010}/SSB_{MSY} vary among jittered runs. Stock Synthesis 3 also includes an excellent data-simulation function to examine whether known parameter values can be reliably estimated, and if not whether there are problematic biases in estimates of management reference points.

4.2.2 Input data

Use of multiple contradictory relative abundance indices

Let us assume that the objective of CPUE standardization is the derivation of an index of population-wide abundance. There is only one trend in real abundance and in this assessment there are several standardized (some unstandardized) indices that each provide a different inference of Pacific-wide stock trend (Figure R4, top panel). There are only two possible conclusions: that (1) all but one of the standardization methods are not operating correctly or (2) all of the standardization methods are not operating correctly. In such cases it is not defensible to fit the model to multiple sets of derived data of which the majority are known to be incorrect. Two problems associated with estimation may occur. The first is an uneven objective surface where different parameter vectors suit the different abundance indices.

This first problem is most common in instances where a small number of contradictory relative abundance indices are included. If many relative abundance indices are included a second problem is that the model may not adequately explain the suite of contradictory relative abundance inferences and simply ignore the relative abundance information altogether. There is evidence that this may be occurring in this assessment. Among sensitivity runs, depletion in 1952 has little bearing on depletion 2010 despite the same relative abundance information (Table R1).

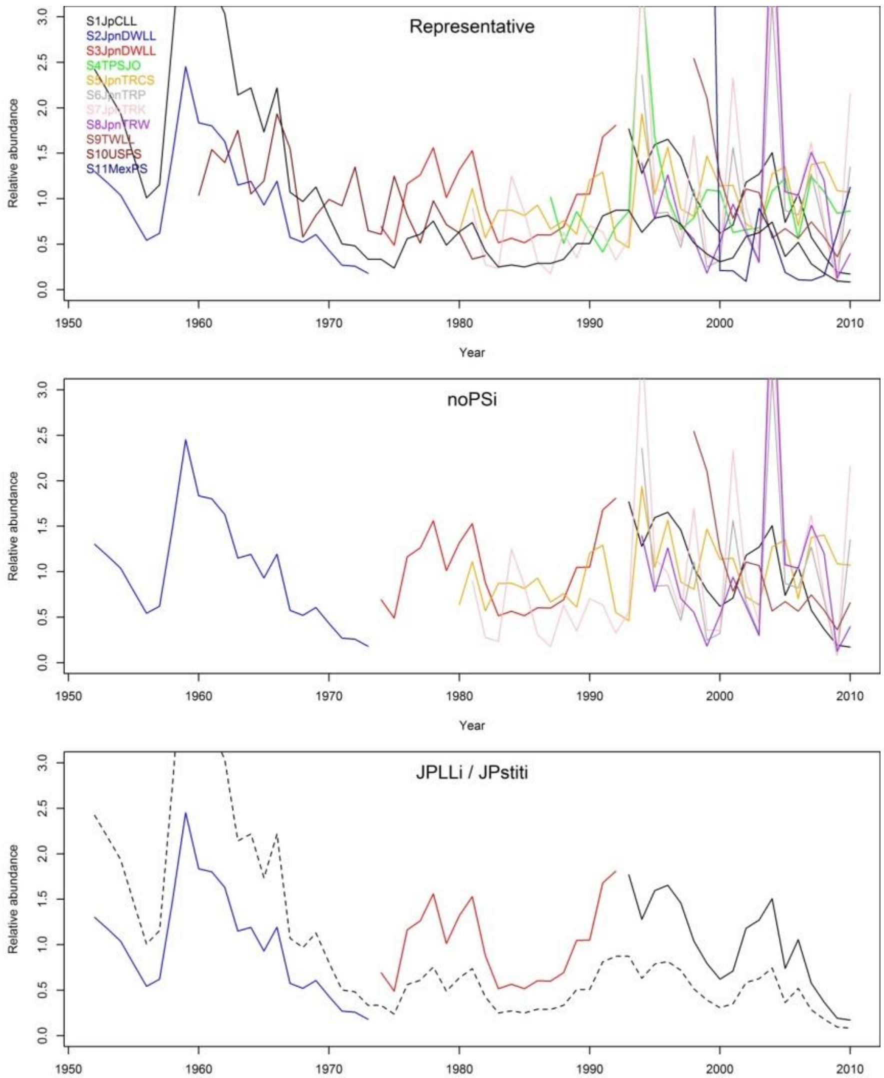

Figure R4. The relative abundance survey indices used in the representative run and three sensitivity analyses that do not use purse seine CPUE series (noPSi), only use Japanese longline CPUE (JPLLi) and 'stitch' together the abundance indices of the Japanese longline fleet (JPstiti, dashed line, bottom panel). For each index the values are normalized to average 1.

Purse-seine CPUE as an index of relative abundance

The use of purse-seine CPUE as an index of abundance is generally not recommended. While the general problems of interpreting commercial catch rate data in terms of abundance are well established there are serious additional issues with purse-seine fishing dynamics. Since these fishing operations target aggregations it is possible for catch rates to remain high while population abundance declines. Note the relatively flat trend in the CPUE index for the important Japanese purse seine fleet (S4TPSJ0, green line top panel Figure R4) relative to the Japanese longline fleet over the same period (Figure R4, bottom panel). This condition, known as hyperstability has been established empirically for purse-seiners (Harley *et al*. 2010) and should preclude the use of these indices in an assessment (*e.g*. S4TPSJ0, Kanaiwa et al., 2012, S10USPS and S11MexPS, Aires-da-Silva *et al*. 2007, Aires-da-Silva and Teo 2012). Note however that removing the purse-seine CPUE had virtually no effect on the estimation (Table R1, Figure R1).

Use of regional relative abundance indices

The inclusion of multiple abundance indices increases the strength of the spatial assumptions of the assessment. It is likely that among years, the regional population density varies even when total abundance remains relatively constant. Fleets with more comprehensive spatial coverage such as the Japanese longline fleet are likely to distribute fishing operations more consistently on the total population and also provide sufficient spatio-temporal coverage to account for time x area interactions in the standardization of CPUE. However, by including an index for a regional fishery (*e.g*. Japanese troll fleets, Ichinokawa *et al*. 2012 or the Taiwanese small scale longline fleet, Hsu and Wang 2012) it must now be assumed that regional trends reflect overall population trends. There is no opportunity to account for inter-annual variation in the regional population density: low regional CPUE is interpreted as low total abundance. Additionally, the inclusion of a regional index may provide extra weight to a particular area of relatively low abundance and may bias the assessment by the location of observations. For these reasons those indices derived from regional fleets should not be used to infer population-wide abundance trends in the representative run of the assessment, particularly since other indices are available.

Use of CPUE that exhibit very high intra-annual variation

Several indices exhibit very rapid changes in inferred abundance among years. For example the Mexican purse seine CPUE series (S11MexPS) is 36 times larger in 1999 than in the year 2000. In 2004 the Japanese Troll fleet (S6JpnTRP) index is 10 times larger than in 2003 and 3.5 times larger than in 2005. Similarly, in 2004 the Japanese troll fleet (S8JpnTRW) is 15 times larger than in 2003 and 4 times larger than in 2005. I acknowledge observation error in these estimates of relative abundance. However, these are not plausible inputs to a population dynamics model that cannot approximate such rapid changes in predicted abundance. In many instances these data-points will be ignored by the assessment. Nonetheless, indices with such fluctuations should be treated with caution and not used in an assessment where alternatives are available. Alternative methods should be investigated to provide more plausible abundance trends.

If one were to investigate the sensitivity of the assessment to the points raised above (regarding multiple indices, the unreliability of purse seine CPUE and the problems with the regional troll gears), a solution may be to include only the three indices for the Japanese longline fleet. This sensitivity analysis (JPLLi) altered estimates of current exploitation level more strongly than a 10% reduction in natural mortality rate across all age classes (M_09). This result underlines the critical role that the survey indices may play in determining reference points.

Use of nominal CPUE

The assessment makes use of nominal purse seine CPUE data (series 4 for the Japanese purse seine fishery) to infer changes in relative abundance. Putting aside the issues of hyperstability in purse-seine CPUE, nominal catch rate data should not be applied in an assessment in circumstances where a range of other standardized indices are available. Nominal purse seine CPUE in particular can be expected to be an unreliable index of relative abundance.

Severing time-series of relative abundance

By dividing-up the relative abundance series of the Japanese longline fleet (S1JpCLL, S2JpnDWLL, S3JpnDWLL) that has good spatio-temporal coverage, long-term inference of depletion is informed to a greater extent by the catch composition data. This is concerning given the fit of the model to these data.

As part of this review, an alternative sensitivity run was undertaken which included the three Japanese longline indices stitched together to investigate the impact on estimated reference points (JPstiti). The departure from the representative run in terms of current stock level and current exploitation level was relatively large and comparable to assuming a natural mortality rate 20% lower across all ages (M_08) (Table R1 and Figure R1). I acknowledge the reasons provided for separating the Japanese longline CPUE indices as presented. However, I would argue that this may be addressed in the standardization phase prior to the assessment. For example, the trip-level CPUE data of the Japanese longline fleet may contain sufficient contrast around the time of the shift to blast-freezer technology to control for this effect and produce a continuous index. Clearly the stitched Japanese longline sensitivity analysis applied in this review is problematic. The question is not about whether such an approach is ideal but whether it is better or worse than relying to a greater extent on the catch composition data to infer stock depletion. This could be addressed by simulation evaluation.

4.2.3 Input parameters

Steepness

The decision to fix steepness at 0.999 is inexplicable. The derivation of steepness is provided by Iwata (2012) who indicates that this value is based on an erroneous calculation:

> *As a result, the estimated values in frequency distribution of steepness are near to one, i.e. within the interval of 0.997-0.999 for both cases of production model and age structured model. So it is appropriate that the steepness value at the coming stock assessment be set as 0.999. However, revised results given in the erratum of Mangel et al. (2010) indicated possible range of steepness of 0.8-1.0. The difference between their result and our result may come from the derivation of <> . Therefore, at the coming stock assessment, the value in the range of 0.8 to 1.0 is recommended for the sensitivity analysis from our results and Mangel's results.*

A better option would be to prescribe the correct prior for steepness alluded to by Iwata (between 0.8 and 1.0, which sounds comparable to that derived by Mangel *et al*. 2010). Stock Synthesis 3 provides suitable options for defining a beta distribution that may offer a reasonable approximation of this prior. The benefits include consistency with life-history characteristics of the stock (which was presumably the aim of the Iwata (2012) analysis) and also that uncertainty may be accounted for in a parameter that is critical to the assessment (as emphasized by Mangel *et al*. 2010).

Natural mortality rate

There are two papers referenced in the assessment document that describe the estimation of natural mortality rate for age 0 and age 1 bluefin tuna (Takeuchi and Takahashi 2006, Polacheck *et al*. 1997, respectively). When the authors refer to Pauly's method for estimating the natural mortality rate of individuals older than age 1, I presume they are referring to the 1980 paper (Pauly 1980). The discussion about natural mortality rate has been covered extensively in earlier meetings and the natural mortality-at-age was heavily revised in this current assessment. In reference to Aires da Silva *et al*. (2008) the Pacific Bluefin Tuna Working Group (PBTWG 2008) documented the following discussion:

> *In the absence of direct estimates of M for PBF beyond age-0 (1+ years), the WG adopted a vector based on assumptions made for southern bluefin tuna (Thunnus maccoyii SBF). This choice should be re-visited and revised considering the differences that exist between the life-history of PBF and SBF. The adoption for PBF of the SBF estimate of M=0.12 yr-1 for the 4+ year old adult fish seems the most problematic. The latter is based on the long life-span of SBF (maximum age of 42) which does not seem to be the case for PBF (maximum observed age of 21 years). In addition, while the mean age at maturity for SBF varies from age 8-12 years, PBF begins to mature at age 3 and are fully mature at age 5. It seems reasonable to assume that such an early investment on reproduction would result in higher natural mortality levels for mature PBF. An alternative M estimate for the adult fish (3+ year) could be taken as the median value (M=0.27) obtained across a large suite of life-history based methods. Estimates of natural mortality for ages 1 and 2 also need to be revised.*
>
> *It was pointed out that the maximum observed age of PBF should not be taken as an estimate of longevity because of the limited sample size of aged PBF (n≈200 for PBF in the age 10+ range) – especially in comparison to SBF. Differences in the exploitation history of PBF and SBF also may have clouded the interpretation of "maximum age" determined from samples taken from recent-year fisheries. However, it was noted that more recently otoliths from PBF have been sampled from some of the largest PBF available in an attempt to gather information from the oldest individuals to produce a more representative growth curve. Consequently, this sampling strategy may have sampled even more than enough of the oldest individuals for determining growth for the population.*
>
> *While PBF differ from SBF in age-at-maturity and growth rate, there are similarities between PBF and eastern Atlantic bluefin tuna (BFT-E) in population vital rates. However, comparisons of PBF made with the BFT-E should be regarded with caution because they are not independent of estimates made for SBF. No direct estimates of adult M (e.g. from tagging experiments) are available for either PBF or BFT-E. For both stocks, the qualitative level of adult M was established by WG consensus after considering knowledge of the biology, comparison with other bluefin species/stocks, fisheries data, and modeling considerations. For BFT-E, the ICCAT WG used the lower bound of the SBF estimates (M=0.10) to quantify BFT-E adult M, while the ISC WG used the mean of the SBF estimates (M=0.12) to quantify PBF adult M.*

Whilst acknowledging the central differences in the maturity and longevity of SBT and PBT referred to above, the use of 0.25yr^{-1} for age 2+ tuna seems high. Coupled with the age 1 estimate of natural mortality rate, this means that under unfished conditions just 5% of age 1 individuals would survive on to an age of 10. This seems counter-intuitive given that despite a long period of overfishing individuals of this age are still routinely observed. The data of the assessment appears to support this point. For example, the Taiwanese longline fleet (fleet 11) mostly catches fish that are over 180 cm (Figure 3-4, page 28). According to the growth curve of Figure 2-3, approximately 90% of these are between the ages of 8 and 16, and approximately half are between the ages of 10 and 14. Under unfished conditions this latter range is less than 5% of individuals. Subject to fishing mortality rate of the kind predicted since 1952, this fleet is assumed to be actively fishing less than 1% of individuals in the population. In my view this is implausible and points to an overly inflated estimate of natural mortality for age 2+ fish.

A more general point is that the modification from a previous age 2+ mortality rate of 0.12yr^{-1} is very large and was not accompanied by sufficient explanation in this assessment document. The detail of this method, the discussion of assumptions and subsequent sensitivity analyses should be proportional to the impact of natural mortality rate on model estimates (which is very high and comparable only to steepness). Teo (2011) illustrates the asymmetry in the risks of assuming *M* is too high and suggests an absolute maximum of 0.25 yr^{-1}. As such, greater explanation is needed as to why this should be the default (representative run) assumption.

4.3. Provide recommendations on improving the treatment of assumptions (e.g. sensitivity analyses) and description of uncertainty in estimates of stock dynamics and management quantities (e.g. reference points).

4.3.1. Sensitivity analyses

Sensitivity to data inputs, data weighting and accounting for structural uncertainty

The assessment document is careful to acknowledge the importance of structural assumptions

> *The influence of these uncertainties on the stock dynamics was assessed by constructing 20 different models, each with alternative data weightings and structural assumptions (Table 1).* (Section 3, page 4)

and proceeds to test the sensitivity of estimated reference points to alternative scenarios for relative abundance indices, data weightings and size composition data:

> *Sensitivity analyses were used to examine the effects of plausible alternative model assumptions or configurations relative to the base case results. The PBFWG examined the sensitivity analyses in Tables 4-2, 4-3, and 4-3-Appendix for this assessment, which were categorized into four themes: 1) CPUE data; 2) effective sample size of Fleet 3 size composition data; 3) fitting different size composition components; and 4) biology.* (Section 4.6.2 page 33).

> *The WG conducted 20 alternative model runs with plausible alternative model configurations and data (see Section 4.6.2, Table 4-2, Table 4-3, and Table 4-3-Appendix), including the Representative Run (Run 2).* (Section 4.5.1. page 38).

Alternative structural assumptions are normally characterized by differences in the population and fishery equations due to a change in the spatio-temporal disaggregation of the model (e.g. an annual model or a two-area model) or different parameterization of stock dynamics (*e.g*. a delay-difference or a surplus production model) that were not tested here. The 20 sensitivity runs of this assessment represent a relatively narrow range of scenarios. For example, a sensitivity run carried out as part of this review which made use of only the Japanese longline CPUE data (JPind) produced reference points (F_{2010}/F_{MSY}, B_{2010}/B_{MSY}) that differed from the representative run to a greater extent than any of the 20 structural runs of the assessment. Much greater differences can be expected by constructing simple spatially disaggregated assessments (*e.g*. a two area, EPO - WPO model) that use electronic tagging data to inform a prior for movement probabilities. Such models may account for a mixed Pacific-wide stock of relatively high viscosity allowing for regional depletions that are likely to substantially alter the inference of stock status and exploitation rates.

Sensitivity to different inputs for natural mortality rate

The results for the sensitivity analysis for natural mortality rate are phrased solely in terms of absolute spawning stock biomass and recruitment in 1990. There may be more appropriate variables for expressing sensitivities that are more relevant to management decision making. It may be the case that, in terms of spawning biomass, "The representative run did not exhibit the same sensitivity to *M* as in past assessment*"* (Section 5.4.5.1, page 40). However, sensitivity in terms of current fishing mortality rate was very high (Table R1 and Figure R1).

Sensitivity of assessment predictions to alternative inputs for steepness

Sensitivity analysis was suggested for steepness but the assessment document includes only the following result:

> *The model, which assumed a lower steepness parameter (h = 0.8), probably did not converge (final gradient is 2860.57). The trends in SSB and recruitment were similar between the Representative Run and the steepness model, which assumed h was 1.0* (Figure 5-26).

It is not surprising that running the model with steepness fixed to 0.999 and 1.000 leads to similar results. Running the representative run under multiple steepness scenarios revealed that the model would not converge for fixed values of steepness below 0.96. Perhaps this is why the previous value of 0.999 was kept for the representative run instead of the prior distribution derived by life-history analysis. The lack of convergence for steepness values below 0.96 is a cause for concern as the life-history analysis of Mangel *et al*. (2010) and Iwata (2012) suggest values as low as 0.8 are credible. An additional cause for concern is the very high sensitivity of estimates of current exploitation rate to fractional changes in the fixed input for steepness. As far as it was possible, this reviewer undertook small changes to the fixed input value (0.99, 0.98, 0.97, 0.96) which led to slightly better overall model fit and rapid increases in the MLEs of current exploitation rate (Figure R1, Table R1).

The correct distribution for the prior on steepness may be a beta distribution (following Mangel *et al*. 2010). I intended to evaluate a beta prior distribution with alpha = 15 and beta =1 parameterization (maximum prior density around 0.98 with a 95% probability interval in the range 0.78 to 0.99) but I could not make sufficient sense of the stock synthesis manual for the inputs of this distribution. Instead a run was undertaken assuming a weakly informed log-normal distribution (mean 0.9, St.Dev, 0.1) bounded between 0.2 and 1. While not necessarily defensible, this was undertaken to evaluate whether in general priors could be assigned for steepness with all other conditions otherwise identical to the representative run. While it was not possible to run the model with fixed steepness below 0.96, assigning a prior appeared to allow convergence below this value. The MLE estimate for steepness given this prior was 0.941 (95% CI: 0.927 and 0.955) and extended the pattern in F/F_{MSY} and B/B_{MSY} of the other steepness sensitivity runs (h99, h98, h97, h96, Figure R1). The global objective function was also smaller at this level than for the other steepness levels (Table R1).

4.3.2. Accounting for uncertainty

Bayesian analysis

It would be desirable to have the Bayesian posterior estimates since they are more straightforward to interpret. It may also be the case that the MLE estimate (the posterior mode) is not a suitable estimate of an expected value if the posterior is strongly skewed (in which case a posterior median is preferable). Several Bayesian outputs would also be useful in diagnosing model overparameterization, in particular MCMC convergence diagnostics and the joint posterior parameter cross-correlation plots. For example, Figure R4 shows the posterior parameter correlation among 24 parameters of the representative run. There appears to be very strong confounding among several parameters associated with recruitment and selectivity, some forming an almost perfect trade-off with one another. Parameter confounding of this type can be expected to pose difficulties for numerical optimization that must find favourable parameter combinations among a ridged multi-dimensional objective surface. MCMC plots of this kind can help to identify areas for model simplification.

4.3.3. Reference points

The assessment report does not include standard MSY reference points making it difficult to understand the status of the stock in terms of a productive biomass and the expected trajectory of the stock given current fishing mortality rate. Two reference points, F/F_{MSY} and B/B_{MSY} are standard in stock assessment and can be used to quickly illustrate the history of the stock and exploitation in a Kobe plot. The assessment document describes the stock as overfished and subject to overfishing but does not quote MSY reference points. How are overfishing and overfished stock status defined for Pacific bluefin tuna?

4.4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.

As presented, the projections do not appear to be consistent with the estimates of the assessment model. Fishing mortality rate in 2002-2004 is estimated to be 2.5 times greater than $F_{0.1}$ and 1.75 times greater than F_{max} (Table 5-4). It seems contradictory therefore that fishing at this exploitation rate should lead to increases in spawning stock biomass:

> *If the fishing mortality is at the 2002-2004 level, SSB was expected to increase, with median SSB in 2030 expected to be around 40,000 mt* (Section 5.5, second paragraph pages 40-41)

This implies higher current and future recruitment, or a significant change in vulnerability, neither of which are described. Assuming that these projections are valid, greater explanation is needed for the reader to understand this result.

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

As in other stock assessment settings (*e.g*. southern bluefin tuna) simulation evaluation (*e.g*. MSE) may be used to design stock assessments and assist decision making. Relevant issues that could be addressed by simulation evaluation include:
(1) the appropriate level of assessment complexity dedicated to fleet dynamics,
(2) the importance of explicitly modelling seasonal population dynamics,
(3) which fleets are likely to generate catch rate data that offer the most reliable information regarding stock-wide abundance (for example based on spatio-temporal coverage),
(4) potential risks of using multiple contradictory relative abundance indices (*e.g*. what are the consequences of down-weighting these data and relying more strongly on size composition data and total catches to inform depletion),
(5) asymmetries in the risk of prescribing positively or negatively biased inputs for natural mortality rate and steepness,
(6) risks associated with assuming complete stock mixing in a spatially structured population with high viscosity (*e.g*. low probability of exchange among EPO and WPO regions),
(7) design of harvest control rules,
(8) the collection of data to improve decision making (value of information analysis)

The standardization of Japanese longline data should be revisited to investigate ways of producing a single continuous index from 1952. Unless all vessels of this fleet simultaneously moved to new technologies and operating procedures there should be sufficient covariate information in trip-level data to explore such a standardization.

Steepness and natural mortality rate largely determine the outputs of the assessment and should continue to be given attention in proportion to their importance. As described above a range of natural mortality rate and steepness scenarios should be simulated in an operating model to investigate problems associated with mis-specifying these inputs in an assessment.

5. Conclusions

Strengths

The assessment was conducted using established software that has been subject to extensive review in other stock assessment settings.

Input files for the assessment were provided allowing the reviewers to examine the model structure, plot data and interactively interrogate the assessment.

The assessment is detailed particularly in its attempt to capture the historical fishing mortality at size of multiple fleets.

The report provides a description of the quality of the various data.

Weaknesses

The assessment makes use of a steepness value that may have been derived erroneously (Iwata 2012).

Model outputs are highly sensitive to alternative credible values for steepness.

The optimization appears unstable at alternative credible values for steepness.

Model outputs are highly sensitive to alternative credible values for natural mortality rate.

The natural mortality rate of individuals age 2+ appears to be high and provides a poorer fit to the data than lower values.

The complexity of the modelled fishing dynamics appears disproportionately high relative to the probable quality of the data and the size of the assumptions regarding spatial population dynamics.

The assessment makes use of indices derived from purse-seine CPUE that may not be suitable.

Nominal catch rate data are used despite the availability of other standardized relative abundance indices.

Regional abundance trend data are used to inform total population trends which relies on stronger spatial assumptions and may bias assessment predictions by regional characteristics.

MSY-based stock status and exploitation metrics are not provided.

Projections seem to be optimistic given other assessment results (*e.g.* fishing mortality rate in the period 2002 -2004 is estimated to be twice F_{max} whilst also projected to result in increases in spawning stock biomass).

The assessment does not appear to fit size composition well and residual errors indicate model misspecification.

Protocols for fixing parameters values and assigning estimation phases are not fully described.

The model appears to be overparameterized: changes to the order in which parameters are estimated leads to convergence on local minima and posterior cross-correlation plots indicate strong parameter confounding.

A very narrow range of sensitivity analyses were conducted. Results for sensitivity runs for natural mortality rate and steepness are not included in sufficient detail (*e.g.* the effect on F_{2010}/F_{MSY}).

6. Recommendations

Greater detail in description of CPUE standardization methods

There are a number of statistical considerations when standardizing CPUE data that were not addressed by any of the standardizations used in this assessment. Where possible, future standardization papers should account for differences in the number of observations among areas (Campbell 2004), different sizes of areas (CPUE is a measure of density) and provide an equation of the form I_t=, where I is the index and the subscript t is the year. These inclusions reassure the reader that the index was calculated correctly, may not be seriously biased and also ensures that the method is reproducible. It would also be desirable for index standardization papers to include a range of sensitivity analyses that could allow several trends to be used in alternative stock assessment runs.

Include a more comprehensive evaluation of structural uncertainty

It would be desirable to see a greater range of alternative model structures in sensitivity runs. For example, by including fewer fleets, removing the seasonal structure or attempting a simple two area (EPO - WPO) population dynamics model. Surplus production, delay-difference and VPA assessments are simple and quick to apply and would provide an interesting context for the results of the more complex statistical catch-at-age analysis of the representative run. These sensitivity runs are likely to better communicate to decision makers the considerable structural uncertainties that may be otherwise masked by presenting the results of a single model structure.

Present results using more conventional reference points for stock status and exploitation rate.

The assessment report describes the Pacific bluefin stock as overfished and subject to overfishing. However, the report does not make use of standard measures of exploitation and stock status. It was relatively difficult to gain an intuition about how far current estimates of stock size and exploitation rate are from those associated with a productive stock size. In addition to reference points such as depletion and fishing mortality rate relative to *Fmax*, it would be desirable for future assessment to include MSY reference points (B/B_{MSY}, F/F_{MSY}). These are often the basis for standard stock assessment outputs such as the Kobe plot in which the model predicted historical stock status (B/B_{MSY}) and exploitation rate (F/F_{MSY}) are plotted against one another (*i.e.* Figure R2).

Identification of clear management objectives including target and limit reference points

More clearly defined management objectives for the Pacific bluefin stock would enable future assessment to be presented in a meaningful framework and support the development of quantitative tools for decision making (*e.g*. MSE).

Greater transparency in the derivation of critical assessment inputs

Relative abundance indices, inputs for steepness and natural mortality rate are generally derived prior to the assessment. Since they are so critical to assessment outputs it makes sense that the methods used to derive inputs for these inputs should be subject to peer-review.

Consider a simpler assessment model

The assessment document does not fully explain the reason for explicitly modelling seasonal population dynamics. I would recommend an annual assessment model and where necessary disaggregation of fleets by season. I would also suggest the alternative fleet disaggregation highlighted in Table R4 (which identifies 8 fleets) as the most detailed fleet structure that should be considered in this analysis. It should be noted that in many settings reliable management decisions can be made using an assessment model containing a very simple fishery structure (2 or 3 fleets of varying vulnerabilities). On top of the annual, coarser fleet approach I would recommend using only the Japanese longline indices and including the indices of other fleets (only those of good spatio-temporal coverage) as sensitivity analyses. I would not recommend using relative abundance indices from regional fisheries (given a single area population dynamics model) and those derived from purse seine CPUE.

Include a prior probability distribution for steepness

I strongly recommend the use of a biologically derived prior for steepness (alluded to by Iwata 2012 and published by Mangel *et al*. 2010) which, in a Bayesian analysis, can be used to account for uncertainty in this critical input. This is particularly important given the relatively large effect of this parameter on estimates of current stock size and exploitation rate.

Provide a much clearer description of the derivation of natural mortality rate

The assessment document includes a cursory reference to the method of Pauly (presumably Pauly 1980) as a basis for deriving natural mortality rate of individuals age 2 and older. This method should be described in greater detail in future assessments and subjected to peer-review.

Conduct a Bayesian analysis

The MCMC run of the representative assessment took 15 hours to conduct 200,000 iterations on my rather dated laptop. This is a small cost relative to the benefits in terms of quantifying uncertainty in outputs (including skew) and understanding parameter confounding. The outputs are also intuitive unlike frequentist confidence intervals which are routinely misinterpreted.

Consider a simpler model with more complex spatial characteristics

It may be possible to use the electronic tagging data that are available to derive priors for movement to support a simple two area (EPO, WPO) population dynamics model that could provide a valuable comparison to the predictions of the spatially aggregated representative run of this assessment.

Avoid using regional abundance indices to infer population-wide stock dynamics.

I recommend fitting the base-case model to a single index of abundance derived from a fleet with good spatial-temporal coverage (in this case this may be the Japanese longline fleet) and use other indices for sensitivity analysis.

Appendix1. Bibliography

Aires-da-Silva, A., Hinton, M.G., and Dreyfus, M. 2007. Standardized catch rates for Pacific bluefin tuna caught by United States and Mexican-flag purse seine fisheries in the Eastern Pacific Ocean (1960-2006). Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 11-18 D.

Aires-da-Silva, A., Maunder, M., Deriso, R., Piner, K., and Lee, H.-H. 2008. An evaluation of the natural mortality schedule assumed in the PBF 2008 stock assessment and proposed changes. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 10-17. December 2008, Ishigaki, Japan. ISC/08/PBF-2/04.

Aires-da-Silva, A. and Teo, S. 2012. Updated standardized catch rates for Pacific bluefin tuna caught by United States and Mexican-flag purse seine fisheries in the eastern Pacific Ocean (1960-2011). Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 31 January-7 February 2012, La Jolla, California, USA. ISC/12/PBFWG-1/18.ecember 2007, Shimizu, Japan. ISC/07/PBFWG-3/01.

Boustany, A.M., Matteson, R., Castleton, M., Farwell, C., Block, B.A. 2010. Movements of Pacific bluefin tuna (*Thunnus orientalis*) in the Eastern North Pacific revealed with archival tags. Progress in Oceanography. 86:94-104.

Campbell, R.A. 2004. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70(2–3): 209–227.

CCSBT, 2011. Report of the sixteenth meeting of the Scientific Committee. Commission for the Conservation of Southern Bluefin Tuna. Bali. http://www.ccsbt.org/userfiles/file/docs_english/meetings/meeting_reports/ccsbt_18/report_of_SC16.pdf.

Fujioka, K., Ichinokawa, M., Oshima, K., and Takeuchi, Y. 2012. Re-estimation of standardized CPUE of Pacific bluefin tuna caught by Japanese offshore longline fisheries operated during 1952-1974. 2012. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 31 January, 7 February 2012, La Jolla, California, USA. ISC/12/PBFWG1/10: 13p.

Fukuda, H., Kanaiwa, M., Tsuruoka, I., and Takeuchi, Y. 2012. A review of the fishery and size data for the purse seine fleet operating in the Japan Sea (Fleet 3).Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 10-17 November 2012, Honolulu, Hawaii, USA. ISC/12/PBFWG-3/03.

Harley, S.J., Myers, R.A., Dunn, A. 2001. Is catch-per-unit effort proportional to abundance? Canadian Journal of Fisheries and Aquatic Science. 58: 1760-1772.

Hsu, H.-C. and Wang, H.-Y. 2012. Abundance index of Pacific Bluefin tuna (Thunnus orientalis) by Taiwanese small-scale longline fleet in the southwestern North Pacific Ocean. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 30 May-6 June 2012, Shimizu, Japan. ISC/12/PBFWG-2/14: 24p.

ICCAT, 2012. Report of the 2012 Atlantic Bluefin tuna stock assessment session. International Commission for the Conservation of Atlantic Tunas. Madrid. Doc. No. SCI-033/2012.

Ichinokawa, M., Oshima, K., and Takeuchi, Y. 2012. Abundance indices of young Pacific bluefin tuna, derived from catch-and-effort data of troll fisheries in various regions of Japan. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 31 January-7 February 2012, La Jolla, California, USA. ISC/12/PBFWG-1/11: 35p.

Iwata, S. 2012. Estimate the frequency distribution of steepness for PBF. ISC/12/PBFWG-3/01

Kanaiwa, M., Shimura, T., Uji, R., Ishihara, Y., and Takeuchi, Y. 2012. CPUE analysis for Japanese purse seine in the Sea of Japan. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 31 January-7 February 2012, La Jolla, California, USA. ISC/12/PBFWG-1/09: 5p.

Mangel, M., Brodziak, J., DiNardo, G., 2010. Reproductive ecology and scientific inference of steepness: a fundamental metric of population dynamics and strategic fisheries management. Fish and Fisheries. 11: 89-104.

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters and mean environmental temperature in 175 fish stocks. J. Cons. Int. Explor. Mer. 39(2): 175-192.

PBTWG. 2008. Report of Pacific Bluefin Tuna Working Group Workshop. Annex 4. International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean. ISC9.

PBFWG. 2010. Report of the Pacific Bluefin Tuna Working Group Workshop, 6-9 July, 2010, Nanaimo, Canada. Annex 7. Report of the Tenth Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 21-26 July 2010, Victoria, British Colombia, Canada.

Polacheck, T., Hearn, W.S., Miller, C., Whitelaw, W., and Stanley, C. 1997. Updated estimates of mortality rates for juvenile SBT from multi-year tagging of cohorts. CCSBTSC/9707/26.

Takeuchi, Y. and Takahashi, M. 2006. Estimation of natural morality of age-0 Pacific bluefin tuna from conventional tagging data. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 31 January-7 February 2006, Shimizu, Japan. ISC/06/PBFWG/07.

Teo, S. 2011. Sensitivity of estimated spawning biomass to natural mortality assumptions in Pacific bluefin tuna assessments. Working paper submitted to the ISC Pacific Bluefin Tuna Working Group Meeting, 6-13 January 2011, Shimizu, Shizuoka, Japan. ISC/11/PBFWG/10: 7p.

Tseng, M-C., Smith, P.J. 2011. Lack of genetic differentiation observed in Pacific bluefin tuna (Thunnus orientalis) from Taiwanese and New Zealand waters using mitochondrial and nuclear DNA markers. Marine and Freshwater Research. 64(3): 198-209.

Appendix 2. CIE statement of work

Attachment A: Statement of Work for Dr. Tom Carruthers

External Independent Peer Review by the Center for Independent Experts

Stock assessment of Pacific bluefin tuna in the North Pacific Ocean

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: Pacific bluefin tuna in the North Pacific Ocean (NPO) are harvested multi-nationally primarily using purse-seine, troll and set net gear. The U.S. catches bluefin mostly in sport fishery and incidentally in the albacore troll and pole-and-line fishery. An assessment of Pacific bluefin tuna in the North Pacific Ocean was conducted by NMFS staff of the Southwest Fisheries Science Center and collaborating scientists from members of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC), within the ISC's Pacific Bluefin Working Group (PBFWG) in 2012. Results of the 2012 assessment indicate that the 2010 biomass is near the lowest since 1952 (22,606 mt) and at about 3.6% of the unfished levels. Fishing mortality for 2007-09 period was high and above all calculated biological reference points; Fishing mortality increased since the last assessment period of 2002-04. Population dynamics were estimated using a fully integrated age-structured model (StockSynthesis v3.23b; SS) fitted to catch, size composition, and catch-per-unit of effort (CPUE) data from 1952 to 2011 provided by PBFWG members. Life history parameters included a length-at-age relationship from otolith-derived ages and natural mortality estimates from a tag-recapture study. The assessment provides the basis for scientific advice on the status of the Pacific bluefin tuna stock and will be the foundation for international management decisions of the Inter-American Tropical Tuna Commission and Western and Central Pacific Fisheries Commission and its Northern Committee, and domestic management decisions by the Pacific Fishery Management Council (PFMC). An independent peer review of the assessment will provide valuable feedback to the PBFWG in conducting future assessments. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have expertise, working knowledge, and recent experience in various subject areas involved in the review: tuna biology; analytical stock assessment, including population dynamics theory, integrated stock assessment models, and estimation of biological reference points; and Stock Synthesis and AD Model Builder. Scientists employed by or have significant interactions with the Inter-American Tropical Tuna Commission (IATTC) and the Western and Central Pacific Fisheries Commission (WCPFC), and the Secretariat of the Pacific Community (SPC), should not be considered as reviewers. Scientists associated with the ISC also should be excluded as reviewers. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review as a "desk" review of the necessary documentation of the current assessment of Pacific bluefin tuna. Therefore, no travel is required.

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

<u>Prior to the Peer Review</u>: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COR, whom forwards this information to the NMFS Project Contacts no later than the date specified in the Schedule of Milestones and Deliverables. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

<u>Pre-review Background Documents</u>: Two weeks before the peer review, the NMFS Project Contact must send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background

information and reports of the current assessment and sensitivity analyses to be peer reviewed. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Documents will include: The PBFWG stock assessment report and some working papers. **Please note that supporting documentation for the review is confidential and reviewers are not to circulate these documents.**

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and CIE Lead Coordinator.** The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements.

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

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

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contacts in advance of the peer review.
2) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than 7 June 2013, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and to Dr. David Die, CIE Regional Coordinator, via email to ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

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

6 May 2013	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
15 May 2013	NMFS Project Contact sends the CIE Reviewers the report and background documents
16 – 31 May 2013	Each reviewer conducts an independent peer review as a desk review
7 June 2013	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
21 June 2013	CIE submits the CIE independent peer review reports to the COR

28 June 2013	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on substitutions. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

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

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

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

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology,
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

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

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI), 22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel - NMFS Project Contact:

Dr. Steve Teo
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037
Steve.Teo@noaa.gov Tel: 1-858-546-7179, Fax: 1-858-546-7003

Dr. Kevin Piner
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037
Kevin.Piner@noaa.gov Tel: 1-858-546-7179, Fax: 1-858-546-7003

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

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background and Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Appendix 3. Typographical, grammatical errors *etc*.

Page 4, second paragraph, first sentence: "The model configuration associated with Run 2 was chosen as the base-case assessment model to determine stock status and provide management advice, acknowledging that while it represents the general conclusions above, the model was unable to reconcile all key data sources (Figure 6)". It is not clear how Figure 6 is relevant to this point (estimated SSB and recruitment).

Page 9, Section 3.2, fourth sentence: "Thus, the PBFWG set the starting year of the models was set to 1952"

Page 27, Section 4.2.3, Equation: unnecessary parenthesis.

Page 36. Figure 5-5 caption. Positive and negative residuals are defined the same way (expected > observed).

Page 48. Figure 5-11. Blank graph.