Review of the 2011 North Pacific Ocean albacore tuna stock assessment

P.L. Cordue

Fisheries consultant

New Zealand

For CIE Independent System for Peer Review

21 December 2011

Executive summary

The Albacore Tuna Working Group completed a stock assessment for North Pacific Ocean albacore tuna in June 2011. The previous assessment in 2006 had used a VPA and growth data from a study in the 1960s. The 2011 assessment used new growth data and a statistical model, implemented in Stock Synthesis 3, which produced substantially different results from the 2006 assessment.

Catch-at-length data and CPUE time series were fitted in the Stock Synthesis model which began in 1966. There were eight CPUE time series in all, with four from surface-based fisheries (troll, pole and line) and four from longline fisheries. Four of the CPUE series covered a period from 1966 or 1972 through to 2009. Constant catchability was assumed for each CPUE time series. The base model fitted to all of the time series but with different assumed levels of precision, chosen by the Working Group, on the basis of their perceived quality as abundance indices. A multinomial likelihood was used for the catch-at-length data in each fishery and, as is often done, effective sample sizes were assumed equal to the number of sampling events (i.e., rather than the number of fish sampled). For the base model, the Working Group modified the multinomial likelihood components by dividing them by 100. Their reason for this choice was that it gave biomass estimates, in preliminary base models, which were of the same scale as those from other tuna assessments (South Pacific albacore, in particular).

The 2011 assessment contains three serious flaws, each of which individually would make the assessment results very unreliable. Collectively, they invalidate the assessment.

The assumption of constant catchability for the time series which extended from 1966 or 1972 up to 2009 is indefensible. No corrections appear to have been made for changes in fleet structure or improved technology and availability of information to increase catch rates. The CPUE time series need to be split into time blocks where there is some chance that fishing efficiency is constant within each time block (and separate catchabilities estimated for each time block).

The CPUE time series for the surface based fisheries show contradictory trends as do the time series for the longline fisheries. It is very poor practice to use contradictory data with the same model run as the model will "average" the contradictory trends. Data sets need to be placed into consistent groups and model runs performed for each group of data sets.

The use of biomass estimates from a South Pacific albacore tuna assessment (in particular) to determine the weighting for the catch-at-length data, in preliminary base models, is indefensible. Data-weightings can be crucial in many assessments, but it is not acceptable to choose the weightings so that the assessment delivers a predetermined result (or falls within a restricted range of results).

The currently defined reference point is an *F*-limit designed to keep biomass within the range of historical estimates. The biomass associated with the *F*-limit, as a proportion of the virgin or un-fished biomass, is driven by the assessment results rather than biological parameters. Conceptually, this is inappropriate as such a reference point provides little or no protection for the stock's spawning potential. Conventional reference points should be adopted.

There is an urgent need for a new stock assessment as there are indications in the data that the stock *may* be under stress (declining proportions of large fish in the catches; peaks in longline catch rates that *may* have been driven by changes in fleet structure and technology, followed by large declines which *may* have been caused by declines in abundance). Considerable data preparation needs to be done before a stock assessment is possible. In particular, defensible abundance indices need to be constructed from an analysis of fisheries data (not just catch and effort data; temporal changes in fleet structure and other factors affecting catch rates need to be included).

Background

The Albacore Tuna Working Group (ALBWG) completed a stock assessment for North Pacific Ocean albacore tuna in June 2011. The previous assessment in 2006 had used a VPA and growth data from a study in the 1960s. The 2011 assessment used new growth data and a forward-projection statistical model which produced substantially different results from the 2006 assessment. An independent review of the assessment was considered essential.

I am one of three CIE reviewers who conducted a desk-top review of the 2011 assessment. This report presents my review findings and recommendations in accordance with the Terms of Reference (ToRs) for the review (Appendix 2, annex 2).

Review Activities

The main assessment document and supporting documents were emailed to the reviewers by the NMFS contact (Dr. Steven Teo, see Appendix 1). In addition to providing the documents, the project contact suggested that he could perhaps answer questions that the reviewers may have.

I read all of the supplied documents in detail prior to sending the project contact some questions with regard to the CPUE indices. I was interested to know if there were any additional documents that "defended" the CPUE time series as indices of abundance. I also asked if project contact had "some insights into why a unit of effort in 1966/1972 is likely to be equivalent to a unit of effort in 2009 in each of the fisheries used to provide indices".

In his reply, the project contact pointed me to various sections in the documents, but these did not offer any explanation as to why the fishing efficiency of the main fleets had been considered to be constant for 40 years.

Summary of findings

The ALBWG are in a transitional phase with regard to their stock assessment methods. The previous assessment, in 2006, used a tuned VPA while the current assessment uses a forward-projection statistical model implemented in SS3. The 2011 assessment also uses new age-at-length data to estimate growth parameters which are very different from those used in the previous assessment. The ALBWG constructed an "audit trail" from the previous assessment to the current assessment via an updated VPA run with the old growth parameters and an SS3 run also using the old growth parameters.

The move from a VPA to a statistical model is appropriate since precise catch-at-age data, for all of the main fisheries, are not available over an extended timeframe. The use of more recent growth data is also justified. However, the SS3 assessment is badly flawed. The assumption of constant catchability for CPUE time series extending from

1966 or 1972 to 2009 is indefensible. Changes in fishing fleets, targeting practices, and the availability of technology and information which could affect catch rates make it highly likely that catchability has changed substantially over a 30 to 40 year period. An additional serious error is that the ALBWG chose the weighting of the catch-at-length data so that the biomass estimates from preliminary base models were similar to the estimates from a South Pacific albacore tuna assessment. Further, time series with contradictory trends were used in the base model, which is very poor practice. There are also several technical issues with regard to the analysis of data and the model structure and parameterization which are a cause for concern.

Each of the TOR are specifically considered below.

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

The ALBWG used a statistical age-structured model to fit to catch-at-length data, conditional age-at-length data, and CPUE indices. This would not be my preferred approach but it is adequate.

The use of a statistical *age*-structured model is appropriate for most species and data sets. It is not the best choice if there are important *length*-based processes occurring in the population or the fisheries. In this case it is normally better to use a statistical length-structured model (e.g., Breen et al. 2003). An age-structured model can be used to fit length data, but it will not do a good job if selectivity processes are length-based because it does not explicitly track the length distribution of the population. Rather, it simply calculates length distributions from the given age-structure of the population, as required (for fitting). Length-based selectivity in an age-structured model is simply age-based selectivity specified by length (i.e., a length-based selectivity is converted to an age-based selectivity, through the growth curve, so that it can be used in the age-based catch equation).

With the extensive catch-at-length data available it would be better to use a length-structured model. Alternatively, the catch-at-length data can be analyzed, outside the model, to produce catch-at-age data (in much the same way that was done for the VPA) which can then be fitted in an age-structured model. This would be my preferred approach as it avoids the problems of temporal and spatial variation in growth that can be hard to incorporate into a stock assessment model. SS3 also provides another alternative through the use of growth-morphs, where the length-structure can be partially tracked by splitting each cohort into a number of growth-morphs (e.g., 3 morphs: slow-growth, average-growth, fast-growth).

2. *Evaluate the assessment model configuration, assumptions, and input parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.*

The base model configuration and assumptions are inappropriate given the available data and the extent to which that data have been analyzed. I cover the main problems with the assessment below, under a number of subject headings.

CPUE

The use of all eight CPUE time series in the base model, with assumed constant catchability for each time series, invalidates the assessment.

The main issue with the use of the CPUE time series is the assumption of constant catchability. It is impossible to think that a unit of effort in 1966 or 1972 is equivalent to a unit of effort in 2009 for any of the main CPUE time series (i.e., S1, S3, S6, S7). Even without changes in technology (GPS, availability of satellite temperature profiles, gear improvements) the changing fleet structure over such a time period could dramatically alter the effectiveness of a unit of effort (e.g., in series S1, CPUE is in units of fish per boat-day; constant catchability assumes that a boat-day in 1966 is equivalent to a boat-day in 2009 which seems very unlikely as the fleet would surely have modernized and average vessel size may have increased).

A second important issue, with regard to the CPUE time series, is that they cannot all be acceptable indices of abundance because they contradict each other (see ALBWG 2011, Figure 7). For example, in the surface-based fisheries, series S5 shows an increasing trend from 1988 to 2002 and then declines steeply through to 2009; while series S1 shows an increasing trend from 1988 to 2009 (albeit that both time series show higher frequency fluctuations). In the longline fisheries, the four time series peak between 1996 and 2001 but they do so in different years and maintain their peaks for different periods; also series S2 declines to very low levels in 2009 compared to the other series.

The ALBWG were aware that the CPUE time series were of varying quality when considered as abundance indices and put most weight onto series S6 and down-weighted other series relative to it. Unfortunately, the assumption of constant catchability is still made for every CPUE series; and series S6 has not been adjusted in any way for changes in technology or fishing practice over the period 1972-2009 (although an initial steep decline in the series from 1966-1971 was deleted due to concerns about changes in fishing practice and fleet structure). The old axiom of “rubbish in, rubbish out” is highly relevant here. It is very bad practice to “throw everything in” and hope that the quality data will somehow provide a signal that is unaffected by the “rubbish”. It is notable that the estimates of biomass are very sensitive to the weightings of the CPUE indices (ALBWG 2011, Fig. 22) and that series S7 (not S6) is the most influential single time series despite being assigned a CV of 40% (ALBWG 2011, Fig. 20).

The fits, of the base model, to the CPUE time series are of varying quality in terms of residual patterns. Series S6 has an excellent fit, S3-S5 and S8 have only minor residual patterns, but S1, S2, and S7 have severe patterns. The large increase in series S7 from 1988 to 2000 (a factor of about 4.4) is inconsistent with the base model which shows an increase in SSB of less than a factor of 2 over a similar period. It seems quite possible that much of the increase in series S7, and the other longline indices during a similar period, has more to do with changes in fleet structure and the availability of technology than it does to do with albacore abundance. The large increase, from 1993 to 1994, in the number of vessels targeting albacore with longline (ALBWG 2011, Figure 6) indicates a substantial change in fleet structure in these fisheries which occurs at the start of the period when there is a large increase in longline catch rates. The cessation of the highseas driftnet fishery in 1993 may perhaps be related to the increase in the number of longline vessels targeting albacore.

Even if the CPUE time series were broken up into periods of relative stability in terms of fleet structure, gear changes, targeting, availability of technology, and the presence/absence of fish attraction devices (FADs), there would still be the issue of whether they were adequately indexing abundance during those time blocks. The potential for hyper-stability in CPUE indices for targeted fisheries is obvious and well known.

The ALBWG needs to do a thorough analysis of catch and effort data from fisheries that could potentially provide abundance indices. They need to consider over which time blocks such fisheries were relatively “stable” (i.e., little change in factors that affect catch rates which cannot be eliminated by standardization) and, within the stock assessment model, estimate separate catchabilities for each time block. Contradictory time series should not be used together in the same model run (Schnute and Hilborn 1993). They should be in separate runs. A CPUE time series should not be used as an abundance index unless it can be defended to some extent. A “defense” must be more than an assertion; it must be backed up by data and analysis. Best practice standardization methods need to be used and the methods and results of such analyses need to be fully documented.

The CPUE documents that were provided for the review were inadequate. The raw data need to be thoroughly analyzed and presented; the reader should be given an understanding of the data and its strengths and weaknesses in terms of providing abundance indices. Changes in fleet composition and targeting practice should be presented and described, as should changes in the spatial distribution of the fishery. Annual sample sizes and the log-book coverage should be tabulated. The estimated effects need to be fully presented. When year-area interactions are estimated and the overall CPUE time series is constructed as a weighted average of the year-area effects, the individual area trends need to be shown and the weights need to be tabulated. A justification for the weighting needs to be given. When year interactions other than area are estimated it must be explained how the CPUE indices were derived (e.g., a year-month interaction was estimated for one series, but no explanation was given as to how the annual trends for each month were combined). Constants should not be added to CPUE to prevent taking the log of zero. If zero catches are considered to be related to abundance (i.e., are not a reporting artifact) then a delta-lognormal model should be used. If the zeroes are unlikely to be related to abundance then an analysis with just the positive catches could be considered. When delta-lognormal models are fitted, the binomial and positive trends need to be shown. Annual percentages of zeroes need to be tabulated.

Catch-at-length

The calculation of the catch-at-length data should involve scaling raw length samples up to the numbers in the sampled catch and then to the numbers in suitably defined strata (probably a post-stratification unless there was a well-defined sampling design). I have not found any documentation on the details of how the catch-at-length was constructed. Perhaps it is all in other documents. I have often found that when the details of the scaling of raw length data are revealed that it has been done in a less than ideal fashion. For each fishery, the sampling design needs to be documented; sample sizes and coverage by year needs to be tabulated; and the scaling procedures need to be detailed.

The base model fits to the catch-at-length data appear to be very bad. This will partly be due to the very low weight given to the catch-at-length data, but will also be due to temporal and spatial variation in growth which cannot be accommodated by a single von Bertalanffy growth curve. Of course, changes in selectivity will also cause some of the residual patterns.

The ALBWG did some time-blocking of selectivity patterns in some of the fisheries. In some cases this was driven by external knowledge of changes in targeting within the defined fishery and in other cases on the basis of poor residual patterns. The former approach is good practice and the latter not so good. Of particular concern was the decision to allow a domed selectivity in fishery F6s1 after 1993 when there were reduced proportions of fish of 100cm or more (ALBWG 2011, Fig 8). The absence of these fish from the catch may be a change in selectivity or it may indicate a reduction of the number of large fish in the population. The change is not abrupt and is not supported by external knowledge of a change in targeting. There is also evidence from the catch-at-length data for fishery F8 that larger fish may be declining in the population since the mid 1990s (ALBWG 2011, Fig 8). And, the last two periods for fishery F2 show a complete absence of large fish (perhaps, in this case, it is a selectivity change).

Initialization

The model was started in 1966 with annual recruitment (age 0) estimated from 1966 to 2009. The model was assumed to be in equilibrium in 1966 at the biomass associated with a constant "historical catch" of approximately 66,000 t annually (being the average of the annual catches from 1952 to 1965 in fisheries F1, F4, and F7). The base model estimate of SSB in 1966 comes in at less than 50% of the virgin level (ALBWG 2011, Table 10).

I find that the "historical catch" feature of Stock Synthesis is a bit unfortunate (it has been there since SS1). The problem is that a constant annual catch over an infinite number of years does not well approximate the usual early catch history (which starts at zero and generally ramps up to a higher level when the fishery is more developed). I don't object too much when the assumed historical catch is at a low level compared to the catches in the developed fishery. However, that is not the case here.

There are two better approaches that can be used. One option is to take the model back to near the beginning of the commercial fishery, or at least to when catches were a lot less (perhaps under 5000 t annually) and assume that biomass is at virgin levels. Another option is to start the model from when catch data are well estimated and there are sufficient composition data to estimate the initial numbers at age in the population. The assumption of an infinite annual historical catch is not needed in either of these cases.

In general, if an historical catch is assumed, then there must at least be some sensitivity tests with higher and lower assumed values (for this assessment, only lower assumed values need to be looked at since the use of the average from 1952 to 1965 will badly over-estimate an appropriate infinite historical annual catch).

Weighting of data
The ALBWG report states: “There is no objective method of establishing weightings (lambda) for different information sources in the SS model”.

SS3 has the option of user-defined multipliers (lambda) for each likelihood component. These are useful for omitting a data source completely (lambda=0) or for doing quick sensitivity tests on down-weighting or up-weighting data sources. They shouldn’t be used in base model runs because then the estimators cease to be maximum likelihood estimators (or, for the multinomial likelihood, the effective sample sizes reported are not actually the ones implicit in the likelihood). The use of lambdas not equal to 1 also makes the calculation of standardized residuals problematic.

Despite the ALBWG’s statement, for several years, there have been objective methods of determining appropriate relative weights of data sets within statistical assessment models (Francis 2011 gives a summary). These methods have all revolved around balancing the input variance assumptions with the variance of the residuals when the data are fitted in the model; they differ only in the method used to achieve the “balance”. Francis (2011) proposed a new method which generally results in much lower weights for composition data than the previous methods. However, even his method does not go so far as the ALBWG who down-weighted the composition data by a factor of 100 after the initial effective sample sizes were specified as the number of tows/sets sampled (i.e., already assuming an effective sample size of only 1 fish per tow/set).

The justification for the severe down-weighting of the composition data is on page 13 of the assessment report: “This choice resulted in the scaling of the estimated quantities within a range that was considered biologically plausible relative to productivity reported in other tuna assessments”. That is, they examined the results from other tuna stock assessments and chose the composition weighting to get the biomass estimates in preliminary base models to a level at which they were comfortable (see ALBWG 2011, Table 4). They particularly considered the South Pacific albacore assessment results where SSB ranged from 253,000–506,000 t, so it is perhaps not surprising that the base model results have a range of SSB from 260,000–504,000 t.

The ALBWG need to read Francis (2011) and consider which method of weighting they will use in future stock assessments. The method used to construct the base model is indefensible.

Presentation of assessment results

No doubt the report simply follows previous practice with regard to reporting assessment results almost exclusively in terms of *F*, relative *F*, and absolute biomass. It would be useful to also present SSB trajectories in terms of percentage of virgin SSB. There is no reference in the text to the fact that the base model is driven down to less than 50% of virgin by the assumed historical catch and declines to about 30% in 1985 and then recovers to almost 60% of the virgin level in 1999 (see Table 8 and do the calculations). The presentation of these results is important when it comes to considering biomass-based biological reference points; and, even *F*-based reference points aimed at the protection of spawning potential.

Reference points

The reference point used, being an *F*-limit which is designed to keep biomass above the average of the 10 lowest historical estimates is conceptually inappropriate. In a stock assessment where the stock is estimated to be lightly exploited, then the objective aims to keep biomass lightly exploited. Conversely, if the stock assessment estimates that there has been heavy exploitation then the objective is to keep the stock heavily exploited. That is, the reference point only makes sense if the stock has been exploited at an appropriate level and the assessment accurately reflects this.

The use of such a reference point makes a mockery of sensitivity runs. When a run indicates that the biomass has always been above, say, 40% of virgin levels, then the biomass associated with the *F*-limit is above 40% of virgin. However, when another run suggests that biomass has been fished down to be below 10% of virgin for the last decade, then the biomass associated with the *F*-limit is below 10% of virgin. As the reference point is designed to keep biomass at historically estimated levels, it moves with each sensitivity run. This is not a good feature; a reference point should be associated with a fixed level of virgin (un-fished) biomass.

There is no need to use such a dangerous reference point. One or more reference points aimed at conserving the spawning potential of the stock should be adopted. I was glad to see that the ALBWG did consider other much more suitable reference points.

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

The ALBWG completed a wide-range of sensitivity runs relative to their base model. They also constructed a sensible “audit trail” from the previous VPA assessment to the current SS3 base model. Unfortunately, the SS3 base model is a very long way from being appropriate and hence the sensitivity runs (generally just altering one aspect of the base model) are only of academic interest. Had the base model been appropriate, then the approach taken by the ALBWG would have almost been adequate. Additional sensitivity runs that should have been done (with an appropriate base run) include lower values for historical catch, and alternative, but realistic, growth parameters (given the new growth data).

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

The use of an equivalent model constructed in R is acceptable for projections. The use of bootstrapping to incorporate parameter uncertainty is possibly adequate but is not ideal. It would be better to have the estimation and projections completed in a Bayesian setting using a single package.

5. *Suggest research priorities to improve our understanding of essential population and fishery dynamics and improvements to the assessment model.*

The research and management priority is to produce a defensible assessment, but much preparation is required before that can be done.

Analysis of fisheries data aimed at providing abundance indices: The data from fisheries which catch (or used to catch) albacore tuna in the North Pacific need to be fully analyzed. It is not enough to simply run the catch and effort data from a fishery through a standardization model and claim to have an abundance index. The history of each fishery needs to be understood in terms of spatial and temporal changes in fleet composition, targeting practice, fishing gear, available technology and information affecting catch rates, by-catch, and catch and effort. Once the context is provided, then standardized analysis can be performed for fisheries that have sufficient temporal and spatial coverage that they may possibly provide abundance indices. Time blocks can be identified for each fishery, within which standardization has some prospect of correcting for changes in fishing efficiency. (Note: separate catchabilities are estimated for each time block in the assessment model.) Catch should be used as the response variable and various effort variables included with the other explanatory variables (i.e., do not assume a linear relationship between catch and a pre-chosen effort variable; let the model choose the effort variable(s) and the form of the relationship).

Review the calculation of the catch-at-length data: The catch-at-length for each fishery should be reviewed to ensure that appropriate methods have been used to scale and stratify the raw data.

Convert the catch-at-length data into catch-at-age data: If cohorts can be tracked across seasons and across years in the length data for some/most of the fisheries then modal analysis can be used to convert catch-at-length to catch-at-age (with a plus-group) for each of the fisheries. This avoids issues with temporal and spatial variation in growth which will cause problems when fitting length data in an age or length structured model.

Cohort slicing based on growth data which may not be representative of fish in the fishery must be avoided.

Choose suitable reference points: Review a variety of reference points aimed at conserving spawning potential for the stock.

Conduct a stock assessment using a statistical forward projection model: There is no defensible assessment for this stock. The ALBWG need to do a new assessment with help from a stock assessment expert(s) who understands good/best practice when using a statistical model.

Two longer-term suggestions:

Validate the ageing method: Ageing by reading otoliths needs to be fully validated.

Collect more growth data: The existing age-at-length data are fairly sparse and show signs of spatial variability. It would be useful to collect additional data and/or age existing data (if there are unread otolith collects).

Conclusions and Recommendations

The transition from VPA methods to statistical models is not easy for a Working Group; experience with VPAs does not provide expertise in statistical modeling. The ALBWG have put a lot of time and effort into the current assessment. They did a good job in terms of providing an "audit trail" from the previous VPA assessment in 2006, which used an old growth curve, to the new SS3 assessment that fitted growth and used new data. However, the new assessment is unusable because of indefensible assumptions with regard to CPUE time series and data-weightings.

CPUE time series for four surface-based fisheries and four longline fisheries were assumed to be abundance indices with constant catchabilities within each time series. This is an indefensible assumption for four of the time series which extended from 1966 or 1972 up to 2009. No corrections appear to have been made for changes in fleet composition or improved technology and availability of information to increase catch rates.

The use of biomass estimates from a South Pacific albacore tuna assessment (in particular) to determine the weighting for the catch-at-length data in preliminary base models is indefensible. Data-weightings can be crucial in many assessments, but it is not acceptable to choose the weightings so that the assessment delivers a predetermined result (or falls within a restricted range of results).

The currently defined reference point is an *F*-limit designed to keep biomass within the range of historical estimates. Conceptually, this is inappropriate as it provides little or no protection for the stock's spawning potential. Conventional reference points should be adopted.

There is an urgent need for a new stock assessment as there are indications in the data that the stock *may* be under stress (declining proportions of large fish in the catches; peaks in longline catch rates that *may* have been driven by changes in fleet structure and technology, followed by large declines which *may* have been caused by declines in abundance). Considerable data preparation needs to be done before a stock assessment is possible. In particular, defensible abundance indices need to be constructed from an analysis of fisheries data (not just catch and effort data; temporal changes in fleet structure and other factors affecting catch rates need to be included). Also, the ALBWG need to be given much more support in terms of technical advice and guidance from one or more experts who understand good/best practice in the use of statistical assessment models.

References

Breen, P.A.; Kim, S.W.; Andrew, N.L. 2003. A length-based Bayesian stock assessment model for the New Zealand abalone (*Haliotis iris*). *Marine and Freshwater Research 54*: 619–634.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci. 68*: 1124–1138.

Schnute, J.T.; Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessment. *Can. J. Fish. Aquat. Sci. 50*: 1916–1923.

Appendix 1: Bibliography of supplied material

Document #	Title	Authors
Not numbered. ALBWG, 2011	Stock Assessment of Albacore Tuna in the North Pacific Ocean in 2011 (For CIE review only). 109 p.	Albacore Working Group
ISC/10/ALBWG-2/07	Standardized CPUE trend and age composition of North Pacific albacore exploited by Taiwanese longline fisheries, 1995-2008	Chiee-Young Chen, Fei-Chi Cheng and Shean-Ya Yeh
ISC/10-3/ALBWG/02	Time series associated with albacore fisheries based in the Northeast Pacific Ocean	Steven L. H. Teo, Hui-Hua Lee, and Suzanne Kohin
ISC/10/ALBWG-3/04	Abundance indices of north Pacific albacore by Japanese longline for SS3 and VPA analyses	Takayuki Matsumoto
ISC/10/ALBWG-3/07	Revision of standardized CPUE for albacore caught by the Japanese pole and line fisheries in the northwestern North Pacific	Kiyofuji, H and Uosaki, K.
ISC/10/ALBWG-3/08	Standardized CPUE and catch-at-age time series of North Pacific albacore exploited by Taiwanese longline fisheries, 1995–2008	Chiee-Young Chen, Fei-Chi Cheng and Shean-Ya Yeh
ISC/11/ALBWG/02	Age and growth of North Pacific albacore (*Thunnus alalunga*)	R. J. David Wells, Suzanne Kohin, Steven L. H. Teo, Owyn E. Snodgrass1, Koji Uosaki
ISC/11/ALBWG/12	Preliminary North Pacific albacore population analysis using VPA-2BOX and future projection using PRO2BOX for 1966 – 2009	Kiyofuji, H., Iwata, S., Kai, M., Ichinokawa, M., Matsumoto, T., Uosaki, K. and Takeuchi, Y.
ALBWG Annex 5, 2006	Report of the Albacore Working Group workshop, November 28 – December 5, 2006	Albacore Working Group
ALBWG Annex 4, 2010	Report of the Albacore Working Group workshop, 12-19 October 2010	Albacore Working Group
ALBWG Annex 9, 2011	Stock assessment of albacore tuna in the North Pacific Ocean in 2011. 4-11 June 2011, 143p.	Albacore Working Group

Appendix 2: Statement of Work for Patrick Cordue

External Independent Peer Review by the Center for Independent Experts

Stock assessment of albacore 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 and manages 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 Technical Representative (COTR), 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: Albacore tuna in the North Pacific Ocean (NPO) are harvested multi-nationally primarily using troll, pole-and-line, and longline gear. The U.S. has major troll and pole-and-line fleets based in Washington, Oregon, and California. An assessment of albacore tuna in the North Pacific Ocean was conducted by staff of the Pacific Islands Fisheries Science Center, 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 Albacore Working Group in FY 2011. Results of the 2011 assessment are more optimistic than the 2006 assessment because fishing mortality in 2006-2008 has decreased relative to 2002-2004 and because of new assumptions regarding growth in the 2011 assessment. Previous assessments used a VPA model with growth curve parameters derived from a study in the 1960s using relatively outdated aging techniques and a limited size range. The current assessment is based on an SS3 model with growth curve parameters derived from a recent study using modern aging techniques and a substantially larger range of fish sizes. The new growth curve, with a lower L_∞, resulted in a higher scaling of biomass and a more positive assessment of stock status.

The assessment provides the basis for scientific advice on the status of the North Pacific albacore stock and will be a key 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 is essential. 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 who are 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 albacore tuna in the NPO. 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.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contacts no later than the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contacts will provide the CIE reviewers with the background documents and reports of the current assessment and sensitivity analyses to be peer reviewed. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

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

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 can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** 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 6 December 2011, 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.

10 November 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
15 November 2011	NMFS Project Contact sends the CIE Reviewers the report and background documents
16–30 November 2011	Each reviewer conducts an independent peer review as a desk review
6 December 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
20 December 2011	CIE submits the CIE independent peer review reports to the COTR

23 December 2011	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and 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 COTR 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 COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) Each CIE report shall 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 COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

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

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

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. Hui-hua Lee
Pacific Islands Fisheries Science Center, 2570 Dole Street, Honolulu, HI 96822
Huihua.Lee@noaa.gov Tel: 1-808-983-5352; Fax: 1-808-983-2902

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

Annex 2: Terms of Reference for the Peer Review

Stock assessment of albacore tuna in the North Pacific Ocean

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

2. Evaluate the assessment model configuration, assumptions, and input parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.

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

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

5. Suggest research priorities to improve our understanding of essential population and fishery dynamics and improvements to the assessment model.

Please note that supporting documentation for the review is confidential, and reviewers are not to circulate these documents.