

**CIE Independent Report**  
**Stock Assessment of North Pacific Swordfish**  
**February, 2010**

Dr. Noel Cadigan  
Fisheries and Oceans Canada,  
Northwest Atlantic Fisheries Center  
St. John's, NL, Canada

## **Executive Summary**

An assessment of swordfish in the North Pacific Ocean was conducted by the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) Billfish Working Group during 2009. This review report was based on the North Pacific swordfish stock assessment review document.

The assessment concluded with a very high degree of confidence that the North Pacific swordfish population biomass was not below Bmsy in 2006, and the population was not experiencing overfishing.

The main conclusion of this review is that some assumptions (i.e. priors) in the assessment model may provide much of the information about Bmsy, Hmsy (optimal exploitation rate), and MSY benchmarks. These assumptions seem too equivocal for stock and harvest advice. Previous assessments concluded that the assessment data were uninformative about the status of the stock relative to MSY benchmarks. This does not seem to have changed in the current assessment, and the reason the assessment could come to very strong conclusions about stock status seems to be related to different, and somewhat untenable, assumptions.

There was no “continuity run” of the last assessment to demonstrate if changes in advice were related to changes in assessment data or changes in the assessment model. The input data were poorly described. The AR is particularly deficient because it provided no discussion of the possibility of changes in fishery efficiency that could cause a change in CPUE catchability, nor any discussion about the accuracy of catch statistics.

## **Background**

Swordfish (*Xiphias gladius*), also known as broadbill swordfish, inhabit a wide region of the Pacific between the latitudes of 50° N and 50° S. They are a highly migratory species with high economic value in both commercial and recreational fisheries. Swordfish in the North Pacific are harvested multi-nationally, primarily using longline gear. The annual total catch has fluctuated around 15,000 mt

since 2001. The U.S. has a major fleet of swordfish longline vessels based in Hawaii and swordfish harpoon and longline vessels in California.

An assessment of swordfish in the North Pacific Ocean was conducted by staff of the Pacific Islands 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). The assessment was conducted within the ISC's Billfish Working Group during 2009. Results of the swordfish assessment are key to international management decisions of the Western and Central Pacific Fisheries Commission and its Northern Committee, and domestic management decisions by the Western Pacific Regional Fishery Management Council (WPFMC) and Pacific Fishery Management Council. An independent peer-review of the assessment is essential.

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract to provide external expertise through the Center for Independent Experts (CIE) to conduct impartial and independent peer reviews of NMFS scientific projects. Terms of Reference (ToRs) for an independent peer review of the stock assessment of North Pacific swordfish were established by the NMFS Contracting Officer's Technical Representative (COTR) and CIE based on the peer review requirements submitted by NMFS Project Contact. Three CIE reviewers were contracted to conduct an impartial and independent peer review in accordance with the Statement of Work (SoW) and ToRs attached in Appendix 2.

## **Role of reviewer**

I performed a 'desk review' of a report on the stock assessment of North Pacific swordfish, in accordance with the SoW and ToRs (see Appendix 2). The review is structured according to the required format and content described in Annex 1.

I received the North Pacific swordfish stock assessment review document on December 17, 2009. This document was the sole background document for the North Pacific swordfish stock assessment review. I requested additional background documents (ISC\_08\_BILLWG\_SS\_04; ISC\_09\_BILLWG\_2\_01) which I received on January 31, 2010. I was instructed that I did not need to consider this material unless I believed that it would be helpful; that is, the material was not listed for review in the statement of work.

My review activities involved critiquing the assessment report, conducting literature reviews of some aspects of the report, and re-calculating some of the benchmarks to demonstrate the impact of process error.

## Summary of findings

There was little description of previous assessments for this stock, including their problems and motivation for the current approach. Previous approaches seem to have been length-based. Usually one changes an assessment model when there is a problem with the previously used approach. The current assessment lacked this context. There was no “continuity run” of the last assessment to demonstrate if changes in advice were related to changes in assessment data or assessment model. The input data were poorly described.

In the rest of this report I refer to the assessment report of Brodziak and Ishimura (2009) as AR.

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

The stock assessment used a power function surplus production model (SPM),

$$B_t = B_{t-1} + RB_{t-1} \left\{ 1 - \left( \frac{B_{t-1}}{K} \right)^S \right\} - C_{t-1} \quad (1)$$

This is a simple assessment model in which population biomass (B) dynamics are modeled for the entire population, and not separately for age and/or length classes. SPM's have a long history and are well studied, although they are generally considered to be less realistic than age-structured models. SPM's are often used for stocks with no or poor age information (e.g. many tropical stocks). SPM's attempt to infer overall population biomass growth rates (R: reproduction plus growth minus natural mortality) and carrying capacity (K) based on simple age-aggregated data on fishery catches and stock size indices. An index is a proportional estimate. SPM's are often used as checks for more realistic age or length based models. However, they sometimes provide even better estimates of some benchmarks than age-structured models (Laloe, 1995). Prager et al. (1996) demonstrated using simulation studies that SPM's could be useful for assessing strongly age-structured populations like North Atlantic swordfish (*Xiphias gladius*), and SPM's have been included in recent assessments for this stock (Prager, 2002). Hence, there is some scientific evidence and precedence for using SPM with a swordfish stock. The method can be adequate and reliable if the data are appropriate.

Data contrast is a key factor that determines the reliability of an SPM. It is well known that SPM's are not reliable unless there is contrast in the time series of catches or indices. In particular, at some period in the time series, catches need to exceed surplus production so that the stock declines. At some other period catches need to be less than surplus production so that the stock increases. This

is sometimes referred to as a ‘two-way trip’. Information about  $R$  in Eq. (1) comes from data at low stock sizes. Information about  $K$  comes from data at high stock sizes but with low catches. The parameters of Eq. (1) are estimated using CPUE indices that are related to stock size via the observation equation

$$I_t = QB_{t-1} \exp(v_t), \quad (2)$$

where  $v_t$  are normal error terms. The  $Q$  parameter tends to be confounded with  $K$  unless there is data at high stock sizes and high catches.

The North Pacific swordfish total catches have some contrast in size over time (see Fig. 1 in AR), particularly during 1951-1964. Since 1970 the catches have been relatively stable. Note that the AR preferred a two-stock scenario with stocks in the western and central Pacific (subarea 1) and in the eastern Pacific (subarea 2); however, they do not present figures for the catches from these subareas. This information was available in a background document (Figures 4.8 and 4.9 in ISC/09/BILLWG-2/01). Most of the total catch came from subarea 1 so the trends in catch in subarea 1 are similar to the trends in total catch. Catches in subarea 2 generally increased during 1960-2000 but have declined steadily since 2001, although with considerable inter-annual variability.

I focus on the Japanese longline CPUE because it is the longest index time series. For the single stock scenario this index was relatively stable during 1951-1980, increased during 1980-1987, declined during 1987-1998, and has been variable with no overall trend during 1998-2006 (see Fig 5.1 in AR). Hence, during the period 1951-1964, when catches approximately doubled, the CPUE index was stable. During the period 1987-1998 when the index declined by approximately 50% the catches varied without trend. The catch and CPUE series do not have the characteristics (i.e. low catch – high index, high catch – low index, high catch – high index) associated with reliable SPM estimates. However, one really needs to try and fit a model to get a better understanding of this. Similar conclusions apply to the subarea1 stock. The situation for the subarea2 stock seems worse. The CPUE index (see Fig. 4.1 in AR) and total catch (see Fig. 4.9 in ISC/09/BILLWG-2/01) have varied similarly except since 2000 when catches declined by approximately 50% but CPUE did not. The subarea2 stock data suggest the stock was not near carrying capacity at the start of the time series. The  $K$  parameter should be difficult to estimate for this stock.

No difficulties in fitting SPM's were reported in the AR, and this seems to conflict with the above conclusion. The authors used a Bayesian approach to parameter estimation and statistical inference. This involved using somewhat informative priors about the values of model parameters, and this is probably why they could estimate all parameter values. This may also be the reason why the three previous stock assessment studies the authors cite (i.e. Kleiber and Yokawa, 2004; Wang et al. 2005, 2007) concluded that there was little contrast in the

North Pacific swordfish CPUE data to estimate stock status relative to biological reference points, whereas in the present assessment the authors were able to do this. I provide additional discussion of this issue under ToR 2.

The SPM included process error. This is good. It is becoming more standard to incorporate process error in population dynamic models to reflect natural variability not accounted for by the model, and thus to increase the relevance of the model. It is necessary to account for process error for realistic (i.e. appropriate) stock projections. However, it has long been recognized that it is very difficult to separate process error from index measurement error in an SPM (e.g. Punt, 2003; de Valpine and Hilborn, 2005). I provide additional discussion of this issue under ToR 2 and 3.

The basis for Bayesian inference is the posterior distribution, which involves high-dimensional integration. The authors used a numerical approach for this (i.e. MCMC). There are convergence issues with this approach. The authors described and appropriately applied diagnostics to demonstrate convergence.

On Pg. 14 the authors suggest that the priors for  $P_t$  are lognormal. This is not correct. Similar to Meyer and Millar (1999; Eq. 9), the conditional prior

$$P_t | P_{t-1}, K, R, S, \sigma^2 \quad (3)$$

is lognormal. The unconditional ‘implied prior’ is complicated but probably not lognormal. Eq. 13 in the AR is also not written correctly. The priors for  $P_t$  are not independent and cannot be written as a product like Eq. 13. In fact, the  $P_t$ ’s will be highly correlated. The conditional prior in Eq. (3) above should be used in Eq. 13 of the AR. These are technical details that do not affect the results of the assessment.

The assessment model may be inadequate because it does not account for errors in catches. The authors provide no information in the AR about the accuracy of catches. However, on pg. 1, they refer to ‘nominal landings’ which is curious terminology and suggests that the catches are not completely accurate. In fact in ISC/09/BILLWG-2/01, it is described that there is uncertainty in the landings statistics and especially how the landings were divided among subareas 1 and 2. This uncertainty is quite typical. The uncertainty may be larger for subarea 2 where the interannual variation in inferred catches is larger (see Fig. 4.9 in ISC/09/BILLWG-2/01). Under ToR 5 I describe an approach that could be used to account for uncertainty in catches, but this is still an open research question for SPM’s and other stock assessment models. A problem is that unaccounted variability in catches will get incorporated into the CPUE measurement error variance and the process error variance, and over-estimating the process error variance can lead to inaccurate stochastic projections and benchmarks (see ToR 3).

***ToR 2: Evaluate the assessment model configuration, assumptions, and input data and 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.***

There was insufficient information presented in the AR to evaluate the input data - both landings and CPUE indices. The AR is particularly deficient because it provides no discussion of the possibility of changes in fishery efficiency that could cause a change in CPUE catchability (Q). It is common to assume that CPUE Q increases over time. Some assessments apply a small percent increase (e.g. some SEDAR stocks), and others (e.g. ICES assessments) limit the length of CPUE time-series because of potential changes in catchability. This issue should have been considered in the assessment, and reported on in the AR. The spatial extent of the fisheries producing the various CPUE indices should also be described, including the fraction of the stock the indices represent. All things being equal, an index that covers a larger fraction of the stock should get more weight in the stock assessment model.

The authors used a power function SPM (Eq. 1 above). I have no specific experience with this model. However, Prager (2002) studied a similar generalized SPM and concluded that it should be applied with skepticism and in conjunction with the more robust Schaefer form. He advised that unless a good external estimate of the model shape was available, the Schaefer model appeared more suitable for routine assessment use on stocks similar to swordfish. He found that estimates of the generalized shape parameter were highly sensitive to outliers. This suggests that the SPM configuration used in the AR may be difficult to estimate and therefore not configured appropriately. The authors should verify that their estimates of the shape parameter (S) are reasonably robust. It makes no sense to me why the posterior means of S in Table 2 (labeled M) of the AR for subareas 1 and 2 are both less than the mean for the single-stock scenario. One explanation consistent with Prager (2002) is that these estimates are responding too much to noise in the data.

The strength and weakness of the Bayesian approach is the use of priors. Most people agree that incorporating objective prior knowledge is good. However, objective prior knowledge is rarely available, and priors are therefore rarely unequivocal. There is also a technical difference in the interpretation of Bayesian probability versus the more commonly used frequentist notion.

The priors used for the SPM K parameter (i.e carrying capacity) were not logical in that the prior means for subareas 1 and 2 did not sum to the prior mean in the single stock scenario. A better defense of the assumed values for the prior means for K is required. The authors state that the K prior mean “values were

chosen to reflect the magnitude of exploitable biomass likely needed to support the observed fishery catches under each scenario”. However, this is not a valid way to set a prior. The prior is supposed to be independent of the data (i.e. Eq. 13 in AR), but the authors used the data somehow to set the prior for K. The authors should have described why, apart from the data, they expect the carrying capacity of the stock as a whole to be about 150 000 mt with 95% CI (credible interval) of roughly 55 000 – 400 000 mt. I suspect that if the working group had to set the prior with no knowledge of the time-series of landings, then they would have chosen a much flatter prior. Using the data to determine the prior is essentially using the data twice, which leads to false precision. I was left wondering why a prior mean of 100 000 mt for the single stock scenario was less appropriate than the chosen value of 150 000 mt, and if the posterior means are sensitive to this assumed value. For reasons described under ToR1, I suspect the data are not that informative about values for K, and I am concerned that the main conclusion of the AR about the status of the stock relative to BMSY (i.e. Figures 6.1-6.3) are not robust to subjective prior specification. I recognize that the posterior means for K (see Table 2 in AR) are somewhat different than the prior means, but always less, which is a concern.

The authors reported on studies by McAllister et al. (2000, 2001) who investigated priors for R based on life-history information and suggested lower mean R’s for Atlantic swordfish (0.4-0.43) than North Pacific swordfish (0.9-1.0). This was mainly due to differences in assumed values for natural mortality. I have little experience with swordfish and cannot criticize the differences. The AR authors assumed a prior R mean of 0.5 with approximate 95% CI (0.19 - 1.3) which is more consistent with the priors suggested by McAllister et al. (2000) for Atlantic swordfish. However, I suspect that the North Pacific swordfish data favors higher values for R. The clue for this suspicion is as follows. The influence of the prior is higher when the fit to the data is poorer. From the results in Table 1, I calculated the total root mean-squared error (RMSE) for all indices:

Stock Scenario	Total RMSE
Single	0.190
Subarea 1	0.202
Subarea 2	0.242

This suggests that the prior for R will have least effect in the single stock scenario, and most effect for the subarea 2 scenario. The highest posterior mean for R occurred in the single stock scenario. This all suggests that the posterior mean for R would be higher if the authors used less informative priors. Superficially it is puzzling why R for the two stocks as a whole (single stock scenario) is larger than R for either stock, but this may be due to the influence of the priors. However, I recognize that RMSE alone does not determine the fit to the data, and that there are variance parameters ( $\sigma^2$  and  $\tau^2$ ) to account for. It would be useful to understand better what the influence of the priors are, and how sensitive stock status determinations are to these priors.

The prior for the production shape parameter had a mean of 1 with 95% CI (0.25 - 4.0). This broad range is inconsistent with the advice in Prager (2002). A prior consistent with Prager (2002) would have a much smaller CV, perhaps 10% giving 95% CI (0.82 - 1.2). The authors should demonstrate if assessment conclusions are sensitive to reasonable alternative values for S, such as S=1?

The prior for Q seems sensible.

The authors provided no information about the levels of process and measurement error they estimated, and if these parameter values were sensitive to assumed priors. The process error in particular is important to estimate for stochastic projections and determining benchmarks (see ToR3). I could not verify the prior means and variances for error variances in the AR. I tried this in R.

It is common to assume that  $B_1 = K$  when estimating SPM's. Usually the argument for this assumption is that prior to the first year in the data time series the catches were low enough that it is reasonable to assume that the population was at carrying capacity. Usually the CPUE is high at the start of the time series but then declines as fishing reduces the biomass. For example, if the fishery exploited at the optimal MSY rate we would expect CPUE to decline by about 50% after several years of fishing.

In the AR the prior mean for  $B_1/K$  set at 0.9. This does not seem appropriate. The Japanese CPUE indices for **subarea 1** do not decline in the first several years of the time series like we would expect if the population were initially near carrying capacity. This may explain the residual pattern during 1952-1963 in the bottom panel of Figure 3.1 in the AR. The model predicted biomass usually declines for the first 12 years as expected because the posterior mean for  $B_1$  (see Figure 6.1) is about 90% of the posterior mean for K (116, see Table 2) and the harvest rate is about 20%. When  $R=0.58$  and  $S=1$  (subarea 1 estimates; see Table 2) then any harvest rate  $> 5.5\%$  will cause the population to decline from  $B_1 = 0.9K$ . The flat trend in the observed CPUE combined with the declining trend in biomass (and hence predicted CPUE) leads to the residual pattern in Figure 3.1. If the harvest rate was constant at 20% in the first part of the time series then  $B_1/K = 65\%$  would produce a flat predicted CPUE and no residual pattern.

For **subarea 2** the prior assumption that  $B_1 = 0.9K$  also does not seem appropriate. Although the harvest rates in the first several years (Fig. 6.2) are low enough to allow for an increasing trend in CPUE (Fig. 4.1),  $B_1/K$  has to be about 50% to get the rate of increase in CPUE over 1955-1970 with a 1% harvest rate based on the values of R, K, and S for subarea 2 in Table 2. I don't understand why predicted biomass decreases in the first 4-5 years when the harvest rate is zero (Fig. 6.2). It should increase. Perhaps the process error cause this.

The assessment model does not account for errors in catches (see comments for ToR 1).

The AR reported on statistical tests for time-trends in residuals, normality of residuals, and constant variance. These tests were not described. I see time-trends in the residuals in Fig. 3.1, but the statistical test was not significant. If the test is for a linear trend then I can understand the result, but this needs to be described in a couple of sentences.

***ToR 3: Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.***

A SPM is convenient in that it provides direct estimates of MSY benchmarks (Bmsy and Fmsy) that may even be more reliable than those from age-structured models (Laloe, 1995). MSY benchmarks are often elusive and usually require knowledge of pre-exploitation stock size and productivity. The AR provided estimates of Bmsy, Hmsy (optimal exploitation rate), and MSY. I suspect that the priors assumed in the SPM provide much of the information about these benchmarks. Some of the priors (i.e. K, B<sub>1</sub>/K) seem too equivocal for stock and harvest advice.

The AR provided expression for Bmsy, Hmsy, and MSY that may not be appropriate for the SPM power model with process error. With this error  $B_t$  is random with distribution  $F_{B,t}$  and the deterministic equilibrium assumption ( $\Delta B_t = 0$ ) has to be replaced by a stochastic equilibrium assumption, such as stationarity ( $F_{B,t+1} = F_{B,t}$ ). The variance of the process error,  $\sigma^2$ , should appear in the calculation of equilibrium reference points and can have important effects. Harvesting according to the deterministic MSY rule is an underoptimized strategy and can lead to strong decreases of the resource (eg. Bousquet et al., 2008). The deterministic Hmsy is incompatible with the assumption of equilibrium: on average, one cannot hope to harvest more than the stochastic MSY. Constant harvesting at the deterministic Hmsy will eventually lead to stock extinction.

Bousquet et al. (2008) showed for a particular type of bounded process error (a “product of beta” distribution) and the Schaefer model that the stochastic Bmsy was

$$B_{MSY} = \frac{K}{2} \left[ 1 - \frac{8\sigma^2}{R(4-R)^2} - \frac{8\sigma^4}{R(4-R)^5} \{3R^3 - 18R^2 + 12R + 32\} + o(\sigma^5) \right]$$

In this equation  $o(x)$  refers to smaller terms, such that

$$\lim_{x \rightarrow 0} \frac{o(x)}{x} = 0.$$

The maximum sustainable yield was

$$MSY = \frac{RK}{4} \left[ 1 - \frac{\sigma^2}{R(1-R/4)^2} + \frac{4\sigma^4}{R^2(4-R)^4} \{R^4 - 4R^3 - 12R^2 + 48R - 16\} + o(\sigma^5) \right]$$

and

$$H_{MSY} = \frac{R}{2} - \frac{2(2-R)\sigma^4}{(4-R)^2} + o(\sigma^3).$$

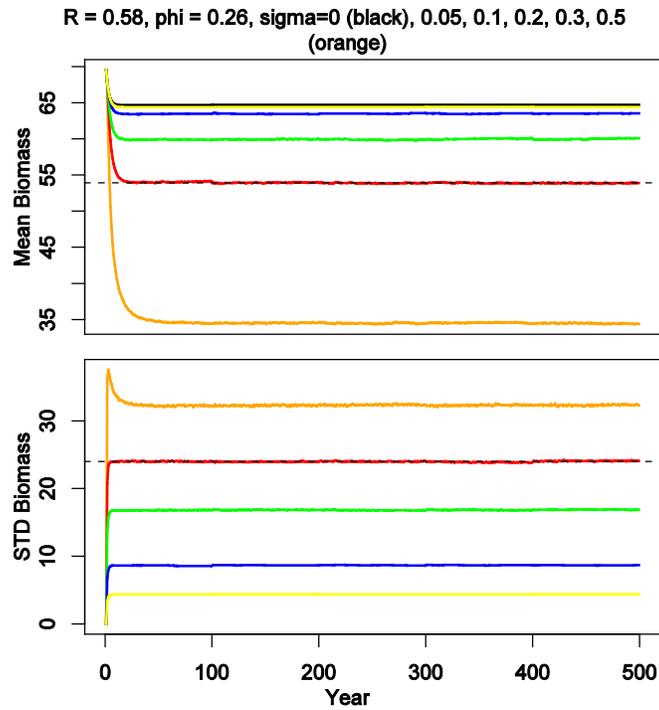
The reference points all equal the usual results when there is no process error; that is,  $\sigma^2 = 0$ . Otherwise they are lower. You can't fish as much or expect as much when there is process error.

In the swordfish SPM the process errors were assumed to be lognormal and the SPM was a generalization of the Schaefer model so the above results do not apply directly; however, they qualitatively indicate the impact of process error, which is a lower Hmsy, MSY, and Bmsy.

I used a simulation procedure to evaluate the MSY benchmarks for the power function SPM based on the values for R, K, and S in Table 2 and a constant harvest rate ( $\phi$ ),  $C_t = \phi B_t$ . The projections are on the "P" scale,

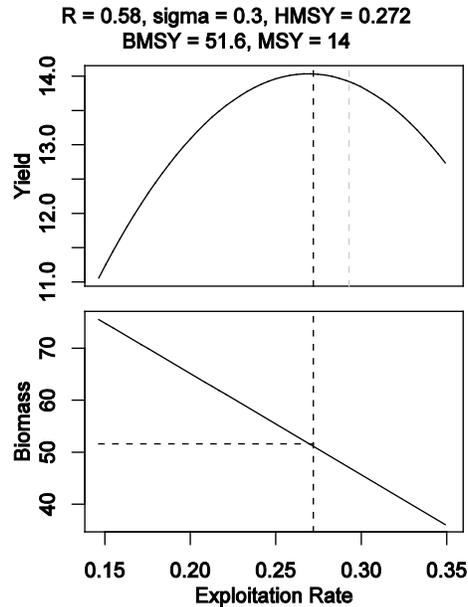
$$\begin{aligned} P_t &= P_{t-1} + RP_{t-1}(1 - P_{t-1}^S) - \phi P_{t-1} \\ &= (1 + R - \phi)P_{t-1} - RP_{t-1}^{1+S} \end{aligned} \quad (4)$$

Note that the AR did not present the posterior means for  $\sigma^2$ , so I computed the benchmarks for a range of possible values. The prior mean for  $\sigma^2$  was 0.025 and I suspect that the posterior mean will not be too different. The first step was to check that the SPM process converged to a stationary distribution. Figure 1 shows results for stochastic projections of the subarea 1 stock based on a constant harvest rate of  $\phi = 0.26$ , which is the Hmsy value in Table 2.



**Figure 1.** Means and standard deviations (STD) from stochastic projection of the subarea 1 stock. The colors indicate different levels of process error dispersion ( $\sigma$ ) which are listed at the top of the panel. A dashed reference line is shown for  $\sigma=0.3$ .

P was fixed at an arbitrary value of 0.6, although the stationary distribution will be independent of this starting value. The results demonstrate that a stationary distribution is achieved (at least to a very good approximation) rapidly. An example of the MSY calculations is shown in Figure 2.  $H_{msy} = 0.272$  which is less than the deterministic result (0.292) obtained using Eq. (4) in the AR.



**Figure 2.** Top panel: The black dashed line indicates the harvest rate that optimizes the stationary distribution mean yield. The grey dashed line indicates the deterministic ( $\sigma=0$ ) result. The bottom panel shown the relationship between the stationary distribution mean biomass and harvest rate. The dashed lines indicate the stochastic Bmsy.

Note that the Hmsy in Table 2 is lower than the result obtained using the posterior means for R and S and Eq. (4) in the AR. This is to be expected because these are means of random variables. The Hmsy results in Figure 2 are based on fixed values for R, S, and  $\sigma^2$ . This could be repeated based on a posterior distribution of values of R, S and  $\sigma^2$  to produce a distribution of values of Hmsy that optimize the stationary distribution mean yield. However, not all values of R, S, and  $\sigma^2$  may produce a stationary distribution so a Bayesian version of Figure 2 requires further investigation.

Non-Bayesian MSY benchmarks for several values of  $\sigma^2$  and the posterior mean values of R and K are given in Table 1 for each stock scenario. If  $\sigma < 0.16$ , which is the prior mean value, then the impact of process error on the MSY benchmarks is not large. However, when  $\sigma \geq 0.3$  the impacts of process error are substantial.

Note that the results in Table 1 are based on the posterior means for R, S, and K. They do not take into account uncertainty in these values. It is interesting to note that the  $\sigma = 0$  results are all higher than the corresponding results in Table 2 of the AR. This suggests that all values in Table 2 may be lower if uncertainty in R, S, and K was accounted for.

**Table 1. MSY benchmarks based on the mean of the stationary distribution for several values of process error,  $\sigma^2$ . The shaded column indicated the results based on the prior mean in the AR.**

Single Stock Scenario								
$\sigma$	0	0.05	0.10	0.16	0.20	0.3	0.4	0.5
MSY	22.4	22.3	22.0	21.4	20.8	18.8	15.9	12.5
Hmsy	0.38	0.38	0.37	0.37	0.37	0.35	0.34	0.30
Bmsy	59.4	59.1	59.1	57.4	56.7	53.5	47.5	41.2
Two-Stock Scenario Subarea 1								
$\sigma$	0	0.05	0.10	0.16	0.20	0.3	0.4	0.5
MSY	17.0	17.0	16.7	16.2	15.7	14.0	11.8	9.2
Hmsy	0.29	0.29	0.29	0.29	0.29	0.27	0.25	0.23
Bmsy	58.2	57.9	57.8	56.8	55.1	51.6	47.1	39.8
Two-Stock Scenario Subarea 2								
$\sigma$	0	0.05	0.10	0.16	0.20	0.3	0.4	0.5
MSY	4.0	4.0	3.9	3.7	3.6	3.1	2.4	1.7
Hmsy	0.16	0.16	0.16	0.16	0.15	0.14	0.13	0.12
Bmsy	25.3	25.2	25.0	23.9	23.5	21.4	18.2	14.3

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

SPM's are appropriate to project future stock status when their simple model assumptions are appropriate. They may not be appropriate when there are strong age or length patterns in population or fishery dynamics and when catches are dominated by recruitment (i.e. a recruitment fishery). SPM's contain an implicit stock-recruitment relationship. They do not model recruitment explicitly like in age-structured models. SPMs are useful for determining benchmarks and some tactical advice, but less useful for short-term strategic advice because they do not have information about the strength of year classes that will recruit to the fishery in the short term.

The AR briefly described the methodology used to project stock status based on recent average fishing mortality rates. It was not described if process error was included in the projections. In Fig. 7.1 for the subarea 1 stock it seemed that the uncertainty decreased with time in the projection which is the opposite of what one expects with process error. I was not clear on how the stochastic harvest rates were generated in the projections. The authors described that the distribution had the same mean as the average of the posterior means for 2004-06, and variance equal to the three year variance in means. This is a very small sample size for determining a variance. Also, it was not clear if the distribution was recalculated for each MCMC sample.

The projections were based on recent average harvest rates but the authors discussed the results in terms of recent average effort. However, harvest rates

and effort are not the same thing. If effort is increased where swordfish are not then harvest rates may not increase much. The issues here are the same as the issues in CPUE standardization.

***ToR 5: Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.***

The sensitivity of important results to priors should be described. The implied priors for time series of biomass and harvest rates, and these quantities relative to Bmsy and Hmsy respectively, should be computed and compared to the posteriors. This could involve figures similar to 6.1-6.3, but with prior and posterior results.

Including more information about the population age and length distributions could improve the evaluation of stock status relative to benchmarks and short term projections. SPM's often try to infer too much (i.e. R and K) based on limited data. If an index of recruitment could be derived from sampling of the size structure of the catches then a very simple stage structured model like CSA (e.g. Mesnil, 2003) could improve the assessment. CSA can compete with more complicated fully age and/or length based assessment models, but CSA is much less data-demanding. It seems like the logical next step beyond SPM. Delay-difference models have also been advocated (e.g. Meyer and Millar, 1998) but they do not utilize stage-information and seem more difficult to estimate than an SPM. They may not offer improvements in situations where SPM is difficult to estimate.

Including measurement error in catches is also important. It would be desirable if external estimates of catch measurement errors were available, but most likely this will not be the case and the measurement error variance will likely need to be inferred in the stock assessment model. It is difficult to separate process error variance and index measurement error in an SPM, and adding catch measurement error is a further complication. However, if one expects the between year variations in harvest rates to be "small" then a possible solution is to treat harvest rates as a random walk or some other type of smooth "drift" model to give some "degrees-of-freedom" to estimate catch measurement error. An example of such a model is:

$$\begin{aligned} \log(P_1) &= \eta_1 \\ \log(P_t) &= \log\left\{(1 + R - \phi_t)P_{t-1} - RP_{t-1}^{1+S}\right\} + \eta_t, t > 1 \\ \phi_1 &= \varphi \\ \log(\phi_t) &= \log(\phi_{t-1}) + \gamma_t, t > 1 \end{aligned}$$

where  $\eta_1, \dots, \eta_T$  and  $\gamma_1, \dots, \gamma_T$  are normal process errors. Stock size indices would be used for estimation, similar to the SPM approach the authors used,

$$\log(I_t) = \log(QKP_t) + v_t$$

$$\log(C_t) = \log(\phi_t KP_t) + \xi_t$$

where  $u_t$  and  $\xi_t$  are index and catch measurement errors.

The accuracy of the assessment and the utility of the short-term projections could be better demonstrated using retrospective analyses of the model. Model results from retrospective runs of the model should be provided.

## Summary of conclusions and recommendations

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

There is some scientific evidence and precedence for using Surplus Production Models (SPM's) with a swordfish stock. The method can be adequate and reliable if the data are appropriate. However, the North Pacific swordfish CPUE and catch time-series for the stock as a whole or for the subarea 1 stock do not have the characteristics associated with reliable SPM estimates. The situation for the subarea2 stock seems worse. The CPUE data suggest the stock was not near carrying capacity at the start of the time series. The K parameter should be difficult to estimate for this stock.

The authors used a Bayesian approach to parameter estimation and statistical inference. This involved using somewhat informative priors about the values of model parameters, and this is probably why they could estimate all the SPM parameters. I conclude some assumptions in the priors seem speculative and not really supported by the data. A sensitivity run with much less informative priors should have been provided. I suspect that results from such a formulation would indicate that the data are uninformative about the status of the stock relative to MSY benchmarks, similar to the previous assessments of Kleiber and Yokawa, (2004) and Wang et al. (2005, 2007)

The assessment model may be inadequate because it does not account for errors in catches.

***ToR 2: Evaluate the assessment model configuration, assumptions, and input data and 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.***

There was insufficient information presented in the AR to evaluate the input data - both landings and CPUE indices. The AR is particularly deficient because it provides no discussion of the possibility of changes in fishery efficiency that could cause a change in CPUE catchability (Q).

The priors used for the SPM K parameter (i.e carrying capacity) were not logical in that the prior means for subareas 1 and 2 did not sum to the prior mean in the single stock scenario. A better defense of the assumed values for the prior means for K is required. The authors used the data somehow to set the prior for K, but assumed the prior was independent of the data. Using the data to determine the prior is essentially using the data twice, which leads to false precision. I am concerned that the main conclusions of the AR about the status of the stocks relative to BMSY (i.e. Figures 6.1-6.3) are not robust to subjective prior specification.

The prior mean for  $B_1/K$  was set at 0.9 in the assessment. This does not seem appropriate. The Japanese CPUE indices for **subarea 1** do not decline in the first several years of the time series like we would expect if the population was initially near carrying capacity. For **subarea 2** the prior assumption that  $B_1 = 0.9K$  also does not seem appropriate.

The SPM configuration used in the AR may be difficult to estimate and therefore not configured appropriately. The authors should verify that their estimates of the production model shape parameter (S) are reasonably robust. The prior for S had a broad range which was inconsistent with the advice in Prager (2002). A prior consistent with Prager (2002) would have a much smaller CV, perhaps 10%.

The prior for Q seems sensible.

The AR provided no information about the levels of process and measurement error estimated, and if these parameter values were sensitive to assumed priors. The process error in particular is important to estimate for stochastic projections and determining benchmarks (see ToR3).

The assessment model does not account for errors in catches.

***ToR 3: Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.***

The priors assumed in the SPM may provide much of the information about Bmsy, Hmsy (optimal exploitation rate), and MSY benchmarks. Some of the priors (i.e. K,  $B_1/K$ ) seem too equivocal for stock and harvest advice.

Harvesting according to the deterministic MSY rule is an underoptimized strategy and can lead to strong decreases of the resource when there is process error. Constant harvesting at the deterministic Hmsy will eventually lead to stock extinction. If the process error  $\sigma < 0.16$ , which is the prior mean value, then the impact of process error on the MSY benchmarks in the AR is not large. However, when  $\sigma \geq 0.3$  the impacts of process error are substantial.

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

SPMs are useful for determining benchmarks and some tactical advice, but less useful for short-term strategic advice because they do not have information about the strength of year classes that will recruit to the fishery in the short term.

It was not described if process error was included in the projections. I was not clear on how the stochastic harvest rates were generated in the projections.

The projections were based on recent average harvest rates but the authors discussed the results in terms of recent average effort. However, harvest rates and effort are not the same thing.

***ToR 5: Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.***

The sensitivity of important results to priors should be described. The implied priors for time series of biomass and harvest rates, and these quantities relative to Bmsy and Hmsy respectively, should be computed and compared to the posteriors.

Including more information about the population age and length distributions could improve the evaluation of stock status relative to benchmarks and short term projections. If an index of recruitment could be derived from sampling of the size structure of the catches then a very simple stage structured model like CSA (e.g. Mesnil, 2003) could improve the assessment.

Including measurement error in catches is also important.

Results from retrospective runs of the model should be provided.

## **Critique of the NMFS review process**

The original contract schedule called for a review from 1-14 October 2009. I did not get the document until December 17, but the review period was extended

from 2 weeks to 2 months to accommodate this delay. Nonetheless this presented some scheduling challenges.

Annex 1 of the SOW was confusing in places. It referred to a summary report which did not exist to my knowledge. Items 2c and 2e in Annex 1 were confusing and probably should be modified for 'desk-reviews'.

## **Appendix 1: Bibliography of materials provided for appointee's involvement**

Brodziak, J. and Ishimura, G. 2009. Stock assessment of North Pacific swordfish (*Xiphias gladius*) in 2009.

Ichinokawa, M. and Brodziak, J. 2008. Stock boundary between possible swordfish stocks in the northwest and southeast Pacific judged from fisheries data of Japanese longliners. ISC/08/BILLWG-SS/04. Working document submitted to the ISC Billfish Working Group Special Session, November 12-14, 2008, Honolulu, Hawaii, USA.

Courtney, D. and Wagatsuma, L. 2009. Input data for a North Pacific Swordfish Stock Assessment using Bayesian Production Models. ISC/09/BILLWG-2/01. Working document submitted to the ISC Billfish Working Group Workshop, 19-26 May 2009, Busan, Korea.

## **Appendix 2: A copy of the CIE Statement of Work**

### **Attachment A: Statement of Work for Dr. Noel Cadigan**

#### **External Independent Peer Review by the Center for Independent Experts**

##### **Stock Assessment of North Pacific Swordfish**

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

**Project Description:** Swordfish in the North Pacific are harvested multi-nationally, primarily using longline gear. The U.S. has a major fleet of swordfish longline vessels based in Hawaii and swordfish harpoon and longline vessels in California. An assessment of swordfish in the North Pacific Ocean will be conducted by staff of the Pacific Islands 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). The assessment will be conducted within the ISC's Billfish Working Group during FY 2009.

Results of the swordfish assessment will be key to international management decisions of the Western and Central Pacific Fisheries Commission and its Northern Committee, and domestic management decisions by the Western Pacific Regional Fishery Management Council (WPFMC) and Pacific Fishery Management Council. 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. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein. The CIE reviewers shall have the expertise, background, and experience to complete an independent peer review in accordance with the SoW and ToRs herein. CIE reviewer expertise shall include fish stock assessment, mathematical modeling, and statistical computing.

**Location of Peer Review:** Each CIE reviewer shall conduct an independent peer review during a desk review of a report on the stock assessment of North Pacific swordfish, whereby no travel shall be required.

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

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 all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review.

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

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

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

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

- 1) The CIE reviewer shall review all background material and reports provided by the NMFS Project Contact as part of the peer review;
- 2) The CIE reviewer shall conduct an independent peer review in accordance with the ToRs (Annex 2);
- 3) No later than 19 February 2010, the 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. The CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in Annex 2;
- 4) The CIE reviewer shall address changes as required by the CIE review in accordance with the schedule of milestones and deliverables.

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

16 December 2009	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
16 December 2009	NMFS Project Contact sends the CIE the background documents
<b>18 December 2009 – 19 February 2010</b>	The reviewer conducts an independent peer review
19 February 2010	The CIE reviewer submits a draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
5 March 2010	CIE submits CIE independent peer review reports to the COTR
19 March 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

**Modifications to the Statement of Work:** Requests to modify this SoW must be made through the Contracting Officer’s Technical Representative (COTR) who submits the modification for approval to the Contracting Officer at least 15 working days prior to

making any permanent substitutions. The Contracting Officer will notify the CIE within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and Terms of Reference (ToR) of the SoW as long as the role and ability of the CIE reviewer to complete the SoW deliverable in accordance with the ToRs and deliverable schedule are not adversely impacted. The SoW and ToRs cannot be changed once the peer review has begun.

**Acceptance of Deliverables:** Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (the CIE independent peer review reports) to the COTR (William Michaels, via [William.Michaels@noaa.gov](mailto: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) the CIE report shall have the format and content in accordance with Annex 1, (2) the CIE report shall address each ToR as specified in Annex 2, (3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

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

#### **Key Personnel:**

William Michaels, Contracting Officer's Technical Representative (COTR)  
NMFS Office of Science and Technology  
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910  
[William.Michaels@noaa.gov](mailto:William.Michaels@noaa.gov) Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator  
Northern Taiga Ventures, Inc.  
10600 SW 131<sup>st</sup> Court, Miami, FL 33186  
[shivlanim@bellsouth.net](mailto:shivlanim@bellsouth.net) Phone: 305-383-4229

#### **NMFS Project Contact:**

Gerald DiNardo, Stock Assessment Program Leader  
Pacific Islands Fisheries Science Center  
2570 Dole Street, Honolulu, HI 96822-2396  
[Gerard.DiNardo@noaa.gov](mailto:Gerard.DiNardo@noaa.gov) Phone: 808-983-5397

Robert Moffitt, Project Contact  
Pacific Islands Fisheries Science Center  
2570 Dole Street, Honolulu, HI 96822-2396  
[Robert.Moffitt@noaa.gov](mailto:Robert.Moffitt@noaa.gov) Phone: 808-983-3742

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

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR, and Conclusions and Recommendations in accordance with the ToRs.
  - a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a detailed summary of findings, conclusions, and recommendations.
  - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
  - c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.
  - d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
  - e. The CIE independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.
3. The reviewer report shall include as separate appendices as follows:
  - Appendix 1: Bibliography of materials provided for review
  - Appendix 2: A copy of the CIE Statement of Work
  - Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

## **Annex 2: Terms of Reference for the Peer Review**

### **Stock Assessment of North Pacific Swordfish**

1. Review of 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 data and 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. Comment on the proposed population benchmarks and management parameters (*e.g.*, *MSY*, *F<sub>msy</sub>*, *B<sub>msy</sub>*, *MSST*, *MFMT*); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.
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 necessary to formulate best management practices.

## **Appendix 3: Membership or other pertinent information from the panel review meeting.**

### **References**

Bousquet, N., Duchesne, T. and Rivest, L.-P. 2008. Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise. *J. Theor. Biol.* 254 65-75.

de Valpine, P., and Hilborn, R. 2005. State-space likelihoods for nonlinear fisheries time-series. *Can. J. Fish. Aquat. Sci.* 62: 1937-1952.

Laloe, F., 1995. Should surplus production models be fishery description tools rather than biological models? *Aquat. Living Resour.* 8, 1-16.

Mesnil, B. 2003. The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data. *Fish. Res.* 63 193-212.

Prager, MH., Goodyear, CP., and Scott, GP. 1996. Application of a surplus production model to a swordfish-like simulated stock with time-changing gear selectivity. *Trans. Am. Fish. Soc.* 125, 729-740.

Prager, MH. 2002. Comparison of logistic and generalized surplus-production models applied to swordfish, *Xiphias gladius*, in the north Atlantic Ocean. *Fish. Res.* 58, 41-57.

Punt, A.E., 2003. Extending production models to include process error in the population dynamics. *Can. J. Fish. Aquat. Sci.* 60, 1217-1228.