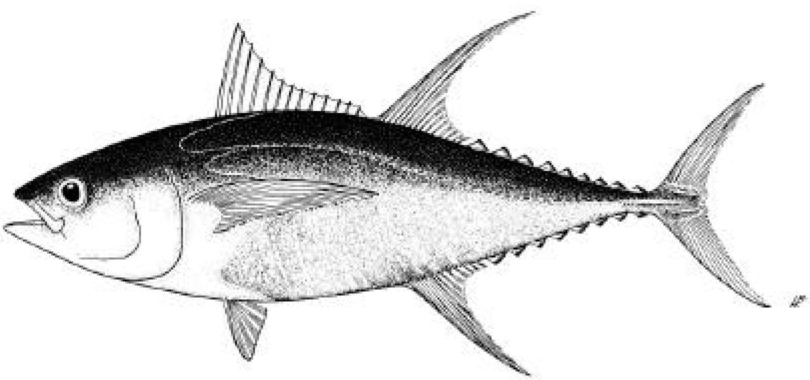
SCTB16 Working Paper

MWG–2

Methodological improvements to the EPO tuna stock assessments



Mark N. Maunder and Shelton J. Harley

Inter-American Tropical Tuna Commission
8604 La Jolla Shores Drive
La Jolla, CA 92037-1508
USA

June 2003

Methodological improvements to the EPO tuna stock assessments

Mark N. Maunder and Shelton J. Harley

Inter-American Tropical Tuna Commission
8604 La Jolla Shores Drive
La Jolla, CA 92037-1508
USA

Tel: (858) 546-7027
Fax: (858) 546-7133
mmaunder@iattc.org

Abstract
This document describes improvements made during 2002-2003 to the stock assessment methodology used by the Inter-American Tropical Tuna Commission to assess tuna in the eastern Pacific Ocean. These improvements include: 1) retrospective analysis to determine years to average catchability for forward projections and yield calculations, 2) cross validation to determine selectivity smoothness parameters, 3) using an analytical formula to re-weight length-frequency sample size, 4) development of a method to allow missing data in environmental indices, 5) calculation of two new management parameters, MSY_{ref} and SBR_{ref}, 6) neural network standardization of CPUE, and 7) likelihood profile approximation to forward projections.

Introduction
The assessments of tunas in the eastern Pacific Ocean (EPO) conducted by the Inter-American Tropical Tuna Commission (IATTC) staff (*e.g*. Maunder and Watters 2001) use A-SCALA (age-structured statistical catch-at-age analysis, Maunder and Watters 2003a), which is based on the MULTIFAN-CL methodology (Fournier *et al*. 1998; Hampton and Fournier 2001) used for assessing tuna stocks in the western and central Pacific Ocean (WCPO) by the Secretariat of the Pacific Community (SPC). Several analyses carried out and presented at previous SCTB meetings have shown that when MULTIFAN-CL is configured to use the same assumptions as A-SCALA the results are essentially the same. With this in mind, it is useful for groups using A-SCALA and MULTIFAN-CL to communicate any improvements in the methodologies so these improvements can be considered for the other methodology.

There have been several improvements made to the assessments of tuna in the EPO during the 2003 assessment process. A majority of these improvements address problems raised at a meeting on diagnostics held in La Jolla on October 2-4 2002, (Harley and Maunder 2003). The outcome from this meeting was 1) a set of diagnostics that should be evaluated regularly, 2) a set of diagnostics that should be evaluated periodically, and 3) a list of specific research questions. These are summarized in Harley and Maunder (2003). This document addresses the research carried out by the IATTC staff to address item 3.

The improvements to the stock assessments discussed in this document are: 1) retrospective analysis to determine years to average catchability for forward projections and yield calculations, 2) cross validation to determine selectivity smoothness parameters, 3) using an analytical formula to re-weight length-frequency sample size, 4) development of a method to allow missing data in environmental indices, 5) calculation of two new management parameters, MSY_{ref} and SBR_{ref}, 6) neural network standardization of CPUE, and 7) likelihood profile approximation to forward projections.

Retrospective analysis to determine years to average catchability

In the previous assessments (Maunder 2002a; Maunder and Harley 2002), two methods were used to determine what fishing mortality or effort and catchability were used in yield calculations and forward projections: 1) fishing mortality averaged over the most recent two years for yield calculations and effort averaged over the most recent two years multiplied by catchability averaged over the most recent two years for forward projections, and 2) effort averaged over the most recent two years multiplied by average catchability over the whole time frame for forward projections. These two methods produced substantially different results for the bigeye tuna assessment (Maunder and Harley 2002). The reason for the difference is that the catchability of bigeye is estimated to have increased for the floating-object fisheries over the last few years. However, using the most recent catchability may not be the best choice because estimates of recent catchability are the most uncertain (Figure 1) and often correlated with recruitment.

We have used retrospective analysis for the 2002 bigeye tuna assessment (Maunder and Harley 2002a) to determine the most appropriate years to average catchability (see Harley and Maunder 2003 for more details on retrospective analysis). Retrospective analysis, for which one year of catch and length-frequency data is removed in consecutive analyses, was carried out while still including effort data for the full time frame of the stock assessment. To include the catchability in the projections, we combined the average effort deviates with the effort used in the projection part of the model. These average effort deviations were calculated over 1) the last two years for which data was available (last two years) and 2) the two years previous to those in 1) above (previous two years). We also included an analysis for which we did not include effort deviates (constant) and a reference based just on the average catch over the estimation period (mean catch). The effort deviation average procedure required that the model be run twice, the first time to estimate the effort deviations used in the averages and the second time to carryout the predictions using these average effort deviations. The retrospective analysis was conducted on an annual basis (one year of data was removed each time), but we present predicted catch for each quarter of the year.

The results indicate that catch is consistently more than the predicted catch for all methods for 1994, 1995, 1998, and 1999 (Figure 2). This is probably due to the increases in catchability during these periods. On average, the last two years performs better than the previous two years, and the constant catchability performs the worst. We also found that correlation plots show that the effort deviates are highly negatively correlated with recruitment for the last year that data are available (Figure 3). This indicates that the last year should not be used to determine catchability.

From this analysis we decided that the best method for yellowfin tuna, which does not show substantial trends in catchability, is for projections to use effort averaged over the last two years (2001 and 2002) and catchability averaged over the two years previous to those (1999 and 2000). For yield calculations, we used average fishing mortality over the two years previous to the last two years (1999 and 2000). We decided that the best method for bigeye tuna, which shows substantial trends in catchability, is for projections to use effort averaged over the last two years (2001 and 2002) and catchability averaged over the second- and third-to-last years (2000 and 2001). For yield calculations we used average fishing mortality over the second- and third-to-last years (2000 and 2001).

Cross validation to determine selectivity smoothness parameters

Numerous assumptions have been used to constrain age-specific selectivity. It is usually not possible to adequately estimate a single selectivity parameter for each age. Therefore, methods are used to limit the number of parameters estimated. The simplest approach is to assume that all individuals above a given age are vulnerable to the fishery and that no individuals below that age are vulnerable to the fishery (knife-edge selectivity). However, the selectivity pattern is rarely so clear cut, so that individuals become gradually more or less vulnerable to the fishery as they age. The logistic curve is frequently used to represent a gradual increase in selectivity as individuals age (*e.g*. Smith and Punt 1998). The assumption that selectivity monotonically increases to an asymptote is implicit in the logistic equation. Many fishing gears may not fully select for older individuals. To include the possibility of decreasing selectivity for the older individuals, other selectivity cures, such as the double normal (Hilborn *et al*. 2000), beta (Punt and Walker 1998), double logistic (Helu *et al*. 2000), and the exponential-logistic (Thompson 1994) have been used.

Haist *et al*. (1999) suggest that these functional forms are too restricting and may be inappropriate for a particular application, leading to biased results. They suggest using separate parameters to represent selectivity for each age, but to constrain the amount that selectivity can change from age to age with smoothness penalties. These penalties avoid overparameterization of the model. The method of Haist *et al* is commonly used in complex statistical catch-at-age or catch-at-length analyses (*e.g*. Fournier *et al*. 1998; Ianelli 2002; Maunder and Watters 2003a). However, the penalties used to determine the smoothness of the selectivity curves are usually specified arbitrarily and can influence the results. For example, the commonly used first, second, and third differences can cause substantial differences in the estimated selectivity curves (Figure 1).

We implement selectivity at age following the method of Fournier *et al*. (1998; see Haist *et al*. 1999 for details). Separate parameters are estimated to represent selectivity for each age, but the amount that selectivity can change from age to age is constrained by smoothness penalties. The constraints are implemented using the difference equation approximation to the first, second, and third derivatives of the selectivity curve. A weighting factor l is added to allow for an increase or decrease in the influence of the selectivity smoothness penalties. The first difference constrains the selectivity curve to be (penalizes it toward being) constant, the second difference constrains it to be linear, and the third difference constrains it to be quadratic. It is likely that selectivity is partly length based, so an additional weighting factor is added to the first difference to apply a higher penalty for ages for which the growth rate is lower and the length distributions are similar between consecutive ages (see Fournier *et al*. 1998 for an alternative method to include length into the selectivity curvature penalty). The parameter y can be modified to determine how the weighting differs as a function of mean length at age. This parameter should be a negative number to provide an inverse relationship between the difference in average length and smoothness of the selectivity curve. If this parameter is set to zero there is no length effect, but the constraint of the first difference is still used. James Ianelli (U.S. NMFS, Seattle, USA, pers. com.,) suggests using the penalties on the logarithm of the selectivity parameters to avoid scale-related problems and improve the stability of the estimation procedure.

First difference

(1) $$l_g^1 \sum_{a=1}^{a=A-1} \left[m_{a+1} - m_a + 0.01\right]^{y_g} \left[-\ln\left(s_{g,a}\right) + \ln\left(s_{g,a+1}\right)\right]^2$$

Second difference

(2) $$l_g^2 \sum_{a=1}^{a=A-2} \left[\ln\left(s_{g,a}\right) - 2\ln\left(s_{g,a+1}\right) + \ln\left(s_{g,a+2}\right)\right]^2$$

Third difference

(3) $$l_g^3 \sum_{a=1}^{a=A-3} \left[-\ln\left(s_{g,a}\right) + 3\ln\left(s_{g,a+1}\right) - 3\ln\left(s_{g,a+2}\right) + \ln\left(s_{g,a+3}\right) \right]^2$$

where

l_g^1, l_g^2, and l_g^3 are the weighting factors for the first, second, and third difference, respectively, for gear type *g*,

y_g, is the length-based weighting factor for gear type *g*,

$s_{g,a}$ is the selectivity to gear *g* for an individual of age *a*,

m_a is the mean length of an individual of age *a*, *A* is the maximum age in the model.

The weighting factors for the selectivity smoothness penalties in the previous assessment were 1, 0, 1, and -1, for the first, second, and third differences, and the length-based penalty, respectively. A weighting factor of 1000 was also applied to a monotonic penalty on gears that are considered to have higher selectivities for the oldest individuals (*e.g.* the southern longline).

We use cross validation to select the appropriate smoothness penalties for the selectivity curves. Cross validation involves using a subset of the data as a test data set. First, the remaining data (the training data set) is used to estimate the model parameters, and then these parameters are used to predict the test data set. The smoothness penalties that provide the predictions that are closest to the test data set are chosen as the best penalties.

For choosing the appropriate smoothness penalties for the selectivity curve, the catch-at-length is used as the data set for prediction. If selectivity is too rough the model may be fitting to noise (due to random sampling error or changes in growth or selectivity) in the catch-at-length training data set, and therefore will not be able to adequately predict catch-at-length in the test data set. Intuitively, this supports using catch-at-length data, rather than other data used in the model (*e.g.* indices of abundance).

Using the bigeye assessment of Maunder and Harley (2002), we randomly chose 20% of the catch-at-length data to be the test data set and used the catch-at-length likelihood function as the measure of closeness of the predicted values to the test data set. The likelihood function is a normal approximation to the multinomial (Fournier *et al.* 1990) with sample size fixed based on assumptions about the data.

The results indicated that weighting factors of 1 on the third difference was appropriate for domed-shaped selectivities (*e.g.* purse-seine fisheries) and a weighting factor of 0.1 on the first difference with a length-based penalty of -1 and a monotonic penalty of 1000 are appropriate for asymptotic selectivity curves (*e.g.* longline fisheries). The results showed that a wide range of smoothness penalties produced similar selectivity curves and estimates of management parameters. In general, smoothness penalties that produced selectivity curves that were not jagged or over-smoothed (*i.e.* too narrow) produced low cross validation scores and similar estimates of management parameters (*e.g.* Table 1).

Analytical formula to re-weight length-frequency sample size

Unfortunately, inclusion of catch-at-age data in stock assessment models can have a large influence on the estimates of trends in abundance (Maunder and Starr 1998). This occurs when the catch-at-age data provide information on biomass that conflicts with the relative abundance trends (*e.g.* catch per unit of effort (CPUE)). Catch-at-age data have been used to estimate fishing mortality rates (*F*) using catch-curve analysis, and therefore should be expected to contain some information on abundance (which is related to catch divided by *F*). However, there are several problems with catch-curve analysis, including assumptions of constant recruitment and constant selectivity. Therefore, several researchers have suggested modifying analyses to ensure that catch-at-age data provide information only on recruitment and selectivity, not abundance.

Fournier and Archibald (1982) suggest that the multinomial is an inappropriate representation of the true estimation error because residuals from a catch-at-age analysis include both observation and process error (Williams and Quinn 1998). The multinomial distribution can represent the observation (sampling) error only under the assumption of random sampling (Williams and Quinn 1998). Sampling designs generally employed to collect fishery-related data generate age-composition estimates that depart from the strict theoretical multinomial probability distribution (Crone and Sampson 1998). The problem with catch-at-age data is the lack of independence in the data, which causes overdispersion. The lack of independence can be due to species schooling or aggregating, producing positive correlations among individuals. Other causes of overdispersion are parameter heterogeneity that is not modelled in the stock assessment and incorrect model structure. The estimators of model parameters often remain unbiased in the presence of overdispersion, but model-based theoretical variances are underestimated (McCullagh and Nelder 1989). However, this is not the case with multiple data types, because the variances affect the weighting among data types and can influence estimates of model parameters.

Fournier and Archibald (1982) suggest that there is variability other than that due to multivariate sampling and that the samples size used in the objective function should be limited to 400. (Note that the more of the variability that is modeled, *e.g.* temporal changes in selectivity, the less the sample size has to be decreased.) Williams and Quinn (1998) found that the effective sample size for an assessment of Pacific herring was substantially less than the actual sample size, and even less than the maximum of 400 suggested by Fournier and Archibald (1982). Williams and Quinn (1998) suggest that since effective sample size can be much smaller than actual sample size, a re-evaluation of the sample size requirements in catch-at-age models is needed. However, there have only been a few studies of the appropriate sample size to use when including catch-at-age data into stock assessment models.

A general method to determine if the appropriate weighting factor (sample size for a multinomial-based likelihood or standard deviation for a normal distribution based likelihood function) is to compare the variance of the residuals (observed minus predicted) to the assumption about the weighting factor. McAllister and Ianelli (1997), used an iterative re-weighting method to determine the effective sample size for catch-at-age data when using a multinomial likelihood function. An initial guess is made of the sample size, the model fit, and then the observed and predicted proportions at age are used in the formula below to calculate the effective sample size. This is then used in the model and the model refit. The process is repeated until the effective sample size does not change.

$$T = \frac{\sum_a \hat{p}_a\left(1-\hat{p}_a\right)}{\sum_a \left(p_a-\hat{p}_a\right)^2}$$

where $\hat{p}_a$ is the predicted proportion and p_a is the observed proportion at age a.

We use McAllister and Ianelli's (1997) method to determine new sample sizes for each set (fishery and time period) of length-frequency data. The original sample size used in the basecase was based on number of wells sampled for the surface gears. For longline we modified the sample size so that the average sample size for the southern longline fishery was equal to the average sample size for the surface fishery that had the maximum average sample size. For example, in the yellowfin assessment, this involved dividing the longline sample size by 25,143 for each length-fequency time-fishery data set.

The reweighting method produces, on average, greater sample sizes for all fisheries (Table 2). The sample size is increased, on average, between about 6 and 18 times for all fisheries except for the northern longline fishery for yellowfin and both longline fisheries for bigeye. This indicates that the purse-seine effective sample size is still less than the number of fish measured (about 50 per well) and that the longline effective sample size is still substantially less than the number of fish measured.

For yellowfin tuna, the estimates of management parameters from the reweighting sensitivity are similar to the basecase, but the confidence intervals are much smaller (Figures 4 and 5). The residuals of the fit to the length-frequency data are much more consistent with the assumptions and sample sizes when using the iterative reweighting (Figures 6 and 7).

For bigeye tuna, the results from the re-weighting sensitivity are quite different to the basecase. Recruitment is much more variable and generally more extreme and the estimates are much more precise. There is a large difference in the biomass trajectories (Figure 8). The strong trend in effort deviates (and therefore catchability) for the southern longline fishery, whose CPUE is standardized and thought to reflect abundance, illustrates the tradeoff between the CPUE and catch-at-length data (Figure 9).

Creation of a method to allow missing data in environmental indices

Previous assessments have used the method of Maunder and Watters (2003b) to integrate the environment-recruitment relationship into the stock assessment model. This method requires that data are available for each time period. Previous assessments could not be initiated before 1980 because environmental data were not available before then. This is an undesired constraint and a method was developed to allow environmental data to be integrated into the analysis even when data are not available for each time period. This was carried out by fitting the estimated recruitment to the predicted recruitment based on the environmental relationship, instead of making recruitment a function of the environmental data. This method is similar to that used by Fournier and Archibald (1982) for integrating a stock-recruitment relationship into the population dynamics model.

The following describes the difference between the Maunder and Watters (2003b) method and the new method.

	Old	New
Model	$R_t = R_0 \exp\left(b\, I_t + e_t \right)$	$R_t = R_0 \exp\left(e_t \right)$
-ln(Like)	$\dfrac{\left(e_t \right)^2}{2s_R^2}$	$\sum_{t=\{\}} \dfrac{\left(\ln\left[R_t \right] - \ln\left[R_0 \exp\left(I_t b \right) \right] \right)^2}{2s_{index}^2} + \ln\left(s_{index} \right) + \sum_{t \neq \{\}} \dfrac{\left(e_t \right)^2}{2s_R^2}$

e_t	$\ln\left(\frac{R_t}{R_0}\right) - b\,I_t$	$\ln\left(\frac{R_t}{R_0}\right)$
Parameters	e_y, b, R_0	e_y, q, s_{index}, R_0, we assume $s_{index} = s_R$

We found that for a model with data for every time period, the new and old method produced the same results.

While investigating the new method we discovered that the error in variables problem applies. In this case it may be informative to compare the results of predictive and calibration regression. Predictive regression (*i.e.* error in the estimates of recruitment) can be implement with

$$\sum_{t=\{\}} \frac{\left(\ln[R_t] - \ln\left[R_0 \exp\left(I_t b\right)\right]\right)^2}{2s_{index}^2} + \ln\left(s_{index}\right) + \sum_{t\neq\{\}} \frac{\left(e_t\right)^2}{2s_R^2}$$

and calibration regression (*i.e.* error in the environmental index) implemented with

$$\sum_{t=\{\}} \frac{\left(I_t - q\ln[R_t / R_0]\right)^2}{2s_{index}^2} + \ln\left(s_{index}\right) + \sum_{t\neq\{\}} \frac{\left(e_t\right)^2}{2s_R^2}.$$

Calculation of two new management parameters: MSY_{ref} and SBR_{ref}

Maunder and Watters (2001) discuss how MSY (maximum sustainable yield) and the SBR (spawning biomass ratio, the spawning biomass divided by the spawning biomass in an unexploited population, SBR_{ref}) at MSY are dependent on the selectivity of the different fisheries and the effort distribution among these fisheries. MSY can be increased or deceased by applying more or less effort to one fishery or another. If the selectivity of the fisheries could be modified at will, there is an optimum yield that can be obtained (Global MSY Beddington and Taylor 1973; Getz 1980; Reed 1980). Maunder (2002b) showed that the optimal yield can be approximated (usually exactly) by applying a full or partial harvest at a single age. He termed this harvest MSY_{ref} and suggested that two thirds of MSY_{ref} may be an appropriate limit reference point (*e.g.* effort allocation and selectivity patterns should produce MSY that is at or above $\frac{2}{3} MSY_{ref}$). The two thirds suggestion was based on analyses in the literature that indicated the best practical selectivity patterns could produce 70-80% of MSY_{ref}, that the yellowfin assessment at the time (Maunder and Watters 2002) estimated that the dolphin fisheries in the EPO would produce about this MSY, and that two thirds is a convenient fraction.

MSY_{ref} is associated with a SBR that may also be an appropriate reference point. SBR_{ref} is not dependent on the selectivity of the gear or the effort allocation among gears. Therefore, SBR_{ref} may be more appropriate than SBR_{MSY} for stocks with multiple fisheries, and should be more precautionary because SBR_{ref} is usually greater than SBR_{MSY}. However, when recruitment is assumed to be constant (*i.e.* no stock recruitment relationship), SBR_{ref} may be dangerous because it is possible that the age that produces MSY_{ref} occurs before the individuals become fully mature. Although it may be possible that a general life history pattern, formed which growth is reduced or natural mortality is increased when individuals become mature, may provide a growth and natural mortality tradeoff after the age at maturity that will produce a SBR_{ref} that is protective of SBR. This is observed for about 90% of the stocks described by Maunder (2002b). SBR_{ref} may be a more appropriate reference point than generally suggested $SBR_{x\%}$ (*e.g.* $SBR_{30\%}$ to $SBR_{50\%}$ see section 5.1) because SBR_{ref} is calculated using the biology of the stock. However, SBR_{ref} may be sensitive to uncertainty in biological parameters such as the steepness of the stock-recruitment relationship, natural mortality, maturity, fecundity, and growth.

For yellowfin tuna in the EPO, MSY_{ref} is estimated to be 417 thousand metric tons, and SBR_{ref} is estimated to be 0.44. If the total effort in the fishery is scaled, without changing the allocation among gears, so that the SBR at equilibrium is equal to SBR_{ref}, the equilibrium yield is estimated to be only 1% less than MSY based on the current effort allocation. This indicates that the SBR_{ref} reference point can be maintained without any substantial loss to the fishery. However, MSY at the current effort allocation is only 61% of MSY_{ref}.

For bigeye tuna, MSY_{ref} is estimated to be 144 thousand metric tons, and SBR_{ref} is estimated to be 0.04. The low SBR_{ref} is a function of the lack of inclusion of a stock-recruitment relationship in the basecase model. This is also consistent with the critical age (18 quarters) being less than the age at which 50% of the females are assumed to be mature. MSY at the current effort allocation is only 47% of MSY_{ref}. If the fishery was only exploited assuming the same selectivity pattern as the southern longline fishery MSY would be 85% of MSY_{ref}.

More research is needed to determine if reference points based on MSY_{ref} and SBR_{ref} are appropriate.

Neural network standardization of CPUE

The effectiveness of longline effort with respect to several tuna species is strongly affected by the fishing depth of the gear, due to the preferences of the species with regard to habitat characteristics (*e.g*. temperature and oxygen levels; Bigelow 2002). Since the mid-1970s, longlines have fished at greater depths in attempts to increase catches of bigeye. Therefore, it is important that standardized longline effort, which is used with catch to provide information on abundance, take into consideration the depth of the longline and the relationship between this depth and the habitat preference of the species.

There are numerous methods that have been used to standardize CPUE. Standardization using generalized linear models (GLMs) is one of the most commonly used methods. GLMs are limited to linear relationships between CPUE and the explanatory variables. An alternative to GLMs are mechanistic models based on our scientific understanding about the relationship between CPUE and the explanatory variables. These models may be complex and nonlinear. One such method that has been used for billfish and tunas is the habitat-based standardization method (HBS) of Hinton and Nakano (1996). GLMs and HBSs are two extremes of the possible methods used to standardize CPUE. The GLM models use little information about the structure of the relationships between CPUE and the explanatory variables and use statistical methods to estimate parameters of the model. The traditional HBS assumes that the relationships between CPUE and the explanatory variables, are known without error and use the current understanding and data to develop the HBS. Hinton *et al*. (2001) developed a statistical HBS that allows the parameters of the sub-models used in the HBS to be updated based fitting to the observed catch and effort data.

One limitation of the statistical HBS is that the structural form of the HBS is fixed, based on scientific understanding. Often this understanding is based on other species or on the same species in different oceans. It is possible that the structure of the HBS model is incorrect, which may cause bias in the estimated year effects that are used in stock assessment models. Therefore, it would be beneficial to develop a model that has a flexible structure that can be estimated from the data.

A neural network allows a very flexible structure to the relationship between dependent and independent variables, and the data are used to estimate the relationship. Neural networks may be viewed as black boxes that take in explanatory variables and output predictions, without providing the ability to easily interpret the relationship between them. If the year effect is all that is desired from the analysis, then as long as the neural network provides good estimates of the year effect, neural networks could be used to standardized CPUE. Unfortunately, because of the difficulty of interpreting the explanatory variables, a standard neural network that takes the year as an input variable cannot be used. We have develop a method based on integrating a neural network with GLM type categorical variables to standardize CPUE data. The year effect is included as a categorical variable, and can be used as an index of relative abundance.

Analyses using several different methods to standardize CPUE (habitat-based methods, statistical habitat-based methods, GLMs, and neural networks) have indicated that neural networks performed as well as or better than other methods based on cross-validation.

The variables included in the neural network were hooks per basket (a measure of depth), latitude, longitude, and the water temperature and oxygen level at a series of depths. Only Japanese catch and effort data are used in the CPUE analysis, because they include information on the number of hooks per basket, provide the only consistent large area coverage of the distribution of yellowfin, and represent the majority of the effort. The effort data used in the stock assessment models are calculated by dividing the total catch for a fishery and time period by the corresponding CPUE (year effect).

The neural network-standardized CPUE produced different trends in biomass for the bigeye tuna assessment (Figure 10), but not for the yellowfin tuna assessment.

Likelihood profile approximation to forward projections

In previous assessments of tunas in the EPO, simulation studies, using the method described by Maunder and Watters (2001), have been conducted to gain further understanding of how, in the future, hypothetical changes in the amount of fishing effort exerted by the surface fleet might simultaneously affect the stocks of tunas in the EPO and the catches of tunas by the various fisheries. This method does not take parameter uncertainty into consideration. It considers only uncertainty about future recruitment. A substantial part of the total uncertainty in predicting future events is caused by uncertainty in the estimates of the model parameters and current status. This uncertainty should be considered in any forward projections. Unfortunately, the appropriate methods are often not applicable to models as large and computationally intense as the A-SCALA and MULTIFAN-CL stock assessment methods. Therefore, we have used a normal approximation to the likelihood profile that allows for the inclusion of both parameter uncertainty and uncertainty about future recruitment. This method is implemented by extending the assessment model an additional five years with effort data based on the average over 2001 and 2002, by quarter (or combining effort with the estimated effort deviates to include more recent estimates of catchability). No catch or length-frequency data are included for these five years. The recruitments for the five years are estimated as in the assessment model with a lognormal penalty with a standard deviation of 0.6. Normal approximations to the likelihood profile are generated for SBR, surface catch, and longline catch.

In general, the estimates from the normal approximation to the likelihood profile are the same as the estimates using the previous method (Figures 11 and 12). The difference occurs in the confidence intervals, which are much larger for the likelihood profile method, particularly for the first year of the projections. These estimates of the confidence intervals are more realistic because they include parameter uncertainty.

Discussion

We present several improvements to the assessments of tunas in the EPO. Many of these improvements are based on initial analyses, and further analyses may be necessary to ensure that the best available methods are used. For example, the confidence intervals are unrealistically small when using the iterative reweighting to determine length-frequency sample size. We hope that this document will provide a starting point for these analyses.

Acknowledgements

Participants of the IATTC scientific review and of the IATTC meeting on diagnostics provided useful comments and advice regarding the improvements to the tuna assessments in the EPO. William Bayliff provided editorial comments on the manuscript.

References

Beddington, J.R. and Taylor, D.B. 1973. Optimum age specific harvesting of a population. Biometrics 29: 801-809.

Bigelow, K.A., Hampton, J., and Miyabe, N. 2002. Application of a habitat-based model to estimate effective longline fishing effort and relative abundance of Pacific bigeye tuna (Thunnus obesus). Fisheries Oceanography 11: 143-155.

Crone, P.R. and Sampson, D.B. 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In: Fishery Stock Assessment Models (Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st Century, October 8-11, 1997, Anchorage, Alaska) F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.J. Schweigert, P.J. Sullivan, and C.I. Zhang eds. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks, pp. 355-370.

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

Fournier, D.A., Hampton, J. and Sibert, J.R. 1998. MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55: 2105-2116.

Fournier, D.A., Sibert, J.R., Majkowski, J. and Hampton, J. 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age-composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (Thunnus maccoyii). Canadian Journal of Fisheries and Aquatic Science 47: 301-317.

Getz, W.M. 1980. The ultimate sustainable yield problem in nonlinear age structured populations. Mathematical Bioscience. 48: 279-292.

Haist, V., Fournier, D., and Saunders, M.W. 1999. Reconstruction of B.C. sablefish stocks, 1966-1998, and catch projections for 1999, using an integrated catch-age mark-recapture model with area and depth movement. Canadian Stock Assessment Secretariat Research Document 99/79 46 pp.

Hampton, J., and Fournier, D.A. 2001. A spatially-disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Fresh. Res. 52: 937-963.

Harley, S.J. and Maunder, M.N. 2003. Recommended diagnostics for large statistical stock assessment models. SCTB16 MWGXX.

Helu, S.L., Sampson, D.B., and Yin, Y. 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Can. J. Fish. Aquat. Sci. 57: 1784-1793.

Hilborn, R., Maunder, M.N., Parma, A., Ernst, B., Paynes, J. and Starr, P.J. 2000. Documentation for a general age-structured Bayesian stock assessment model: code named Coleraine. Fisheries Research Institute, University of Washington, FRI/UW 00/01.

Hinton, M.G. and Nakano, H. 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from Japanese longline fisheries in the Pacific. IATTC Bull. 21: 169-200.

Hinton, M.G., Maunder, M.N., and Uozumi, Y. 2001. Status of striped marlin, Tetrapturus audax, stocks of the eastern-central Pacific. 3rd International Billfish Symposium, Cairns, Australia, 19-23 August, 2001.

Ianelli, J.N. 2002. Simulation analysis testing the robustness of productivity determinations from west coast pacific ocean perch stock assessment data. North American Journal of Fisheries Management, 22: 301-310.

Maunder, M.N. 2002a. Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 3: 47-134.

Maunder, M.N. 2002b. The relationship between fishing methods, fisheries management and the estimation of MSY. Fish and Fisheries, 3: 251-260.

Maunder, M.N. and Harley, S.J. 2002. Status of bigeye tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 3: 201-311.

Maunder, M.N. and Harley, S.J. in preperation. Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 4: xxx-xxx.

Maunder, M.N. and Starr, P.J. 1998. Validating the Hauraki Gulf snapper pre-recruit trawl surveys and temperature recruitment relationship using catch at age analysis with auxiliary information. New Zealand Fisheries Assessment Research Document 98/15 23p.

Maunder, M.N. and Watters, G.M. 2001. Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report 1: 5-86.

Maunder, M.N. and Watters, G.M. 2002. Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 2: 5-90.

Maunder, M.N. and G.M. Watters. 2003a. A-SCALA: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the eastern Pacific Ocean. IATTC Bull., 23: in press.

Maunder, M.N. and Watters, G.M. 2003b. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fishery Bulletin, 101: 89-99.

McAllister, M.K. and Ianelli, J.N. 1997. Bayesian stock assessment using catch-at-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54: 284-300.

McCullagh, P. and Nelder, J.A. 1989. Generalized linear models. Second edition. Chapman and Hall, New York, New York, USA.

Punt, A.E. and Walker, T.I. 1998. Stock assessment and risk analysis for the school shark (*Galeorhinus galeus*) off southern Australia. Marine and Freshwater Research 49: 719-713.

Reed, W.J. 1980. Optimum age-specific harvesting in a nonlinear population model. Biometrics 36: 579-593.

Smith, A.D.M. and Punt. A.E. 1998. Stock assessment of gemfish (*Rexea solandri*) in eastern Australia using maximum likelihood and Bayesian methods. In: *Fishery Stock Assessment Models* (Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st Century, October 8-11, 1997, Anchorage, Alaska) F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.J. Schweigert, P.J. Sullivan, and C.I. Zhang eds. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks, pp. 245-286.

Thompson, G.G. 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. Can. J. Fish. Aquat. Sci. 51: 2654-2664.

Williams, E.H. and Quinn, T.J. II. 1998. A parametric bootstrap of catch-at-age composition using Dirichlet distribution. In: Fishery Stock Assessment Models (Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st Century, October 8-11, 1997, Anchorage, Alaska) F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.J. Schweigert, P.J. Sullivan, and C.I. Zhang eds. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks, pp. 371-384. ?

Table 1. Results from the bigeye tuna application with different weighting factors for the first difference. Bold numbers represent the minimum values for the cross validation (C-V) scores.

	0.01	0.05	0.1	0.5	1	2	10
PS C-V	108.28	**105.77**	105.82	110.01	114.38	123.81	150.04
LL C-V	**0.64**	2.04	2.19	4.25	4.05	5.18	14.44
Total C-V	108.91	**107.81**	108.02	114.26	118.43	128.98	164.49
Scaled C-V	**2.81**	6.60	7.01	12.64	12.14	15.30	40.74
MSY	69808	67923	62809	58845	58043	63046	68083
S_{MSY}/S_0	0.18	0.18	0.19	0.22	0.22	0.28	0.27
S_{cur}	60931	68775	53548	75102	62398	37475	31342
S_{cur}/S_{MSY}	2.44	2.50	2.08	2.15	1.86	0.97	0.83
F_{scale}	1.61	1.63	1.38	1.33	1.17	0.70	0.64

Table 2. Sample sizes assumed in the EPO yellowfin and bigeye tuna 2003 assessments and the sample sizes calculated using the iterative reweighting procedure.

	Yellowfin			Bigeye		
Fishery	Basecase	Reweighted	Average scaling factor	Basecase	Reweighted	Average scaling factor
1	8.38	41.39	8.72	3.8	46.3	15.8
2	5.54	39.28	14.11	13.8	162.8	14.4
3	12.95	51.77	5.53	12.4	132.3	14.8
4	8.56	63.04	10.64	2	29.1	16
5	28.68	147.61	6.13	7.3	99.4	15.5
6	21.84	85.80	5.57	6.5	58.2	14
7	35.31	287.96	11.08	3.1	41	17.6
8	32.59	247.72	9.23	5.8	190.6	229.9
9	8.45	115.36	17.55	13.8	870.7	106.9
10	11.98	76.99	9.24			
11	4.06	150.10	88.10			
12	35.31	314.03	15.48			

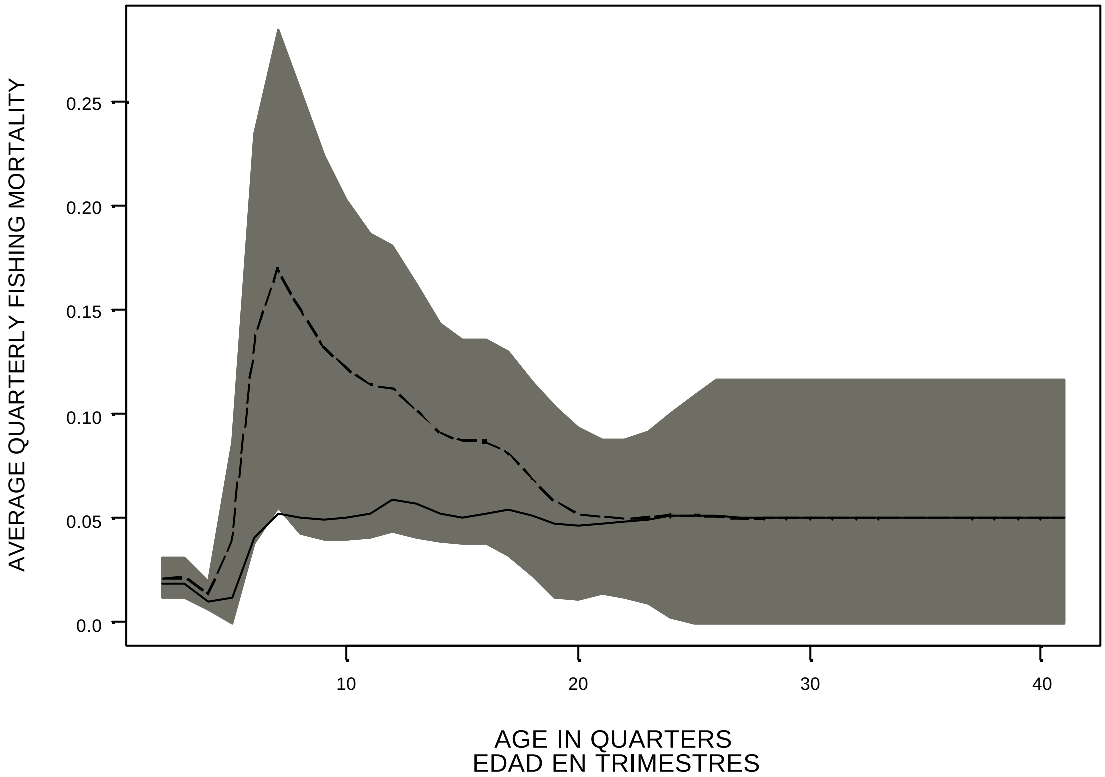


Figure 1. Average quarterly age-specific fishing mortality used in the yield calculations and projections for the EPO bigeye tuna 2002 (Maunder and Harley 2002) basecase (solid line) and sensitivity analysis (dashed line). The basecase is based on average catchability over the whole modeling time period and the sensitivity is based on average catchability over the last two years. The shaded area represents the 95% confidence intervals for the estimated average quarterly age-specific fishing mortality used in the sensitivity analysis.

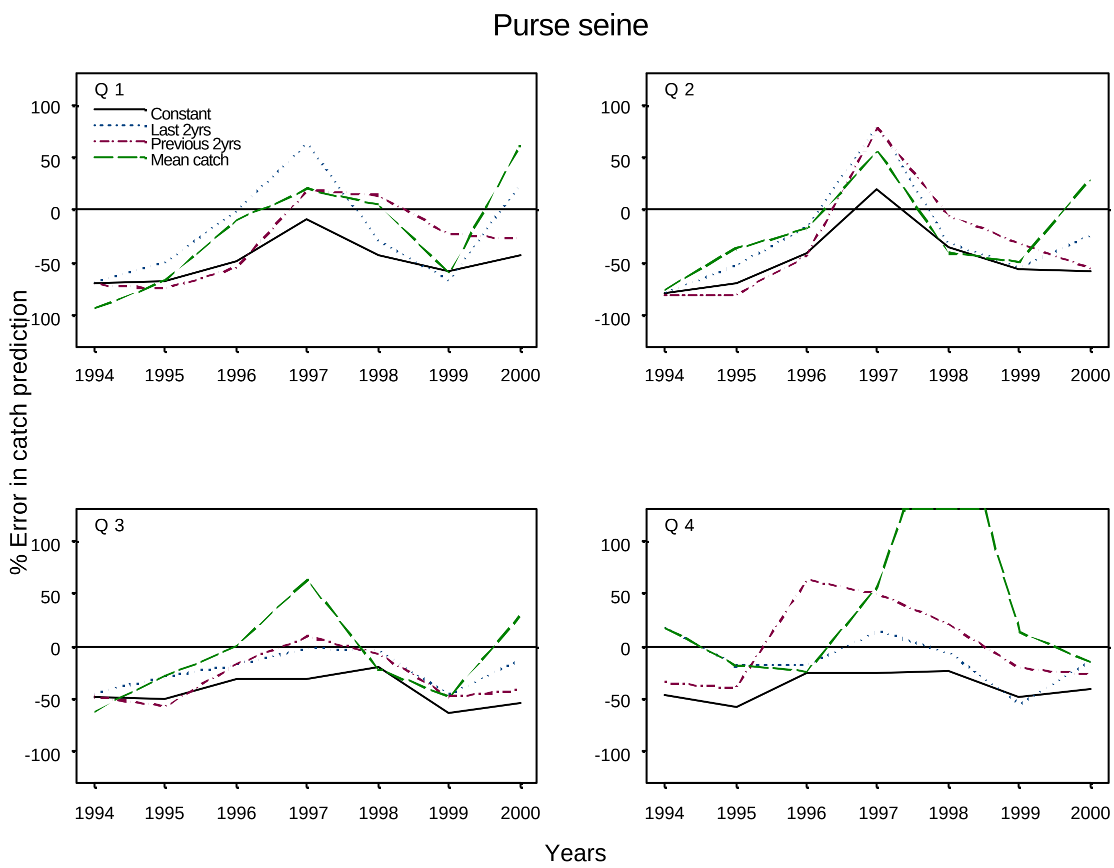


Figure 2. Percentage error in predicted catch by quarter from the retrospective analysis.

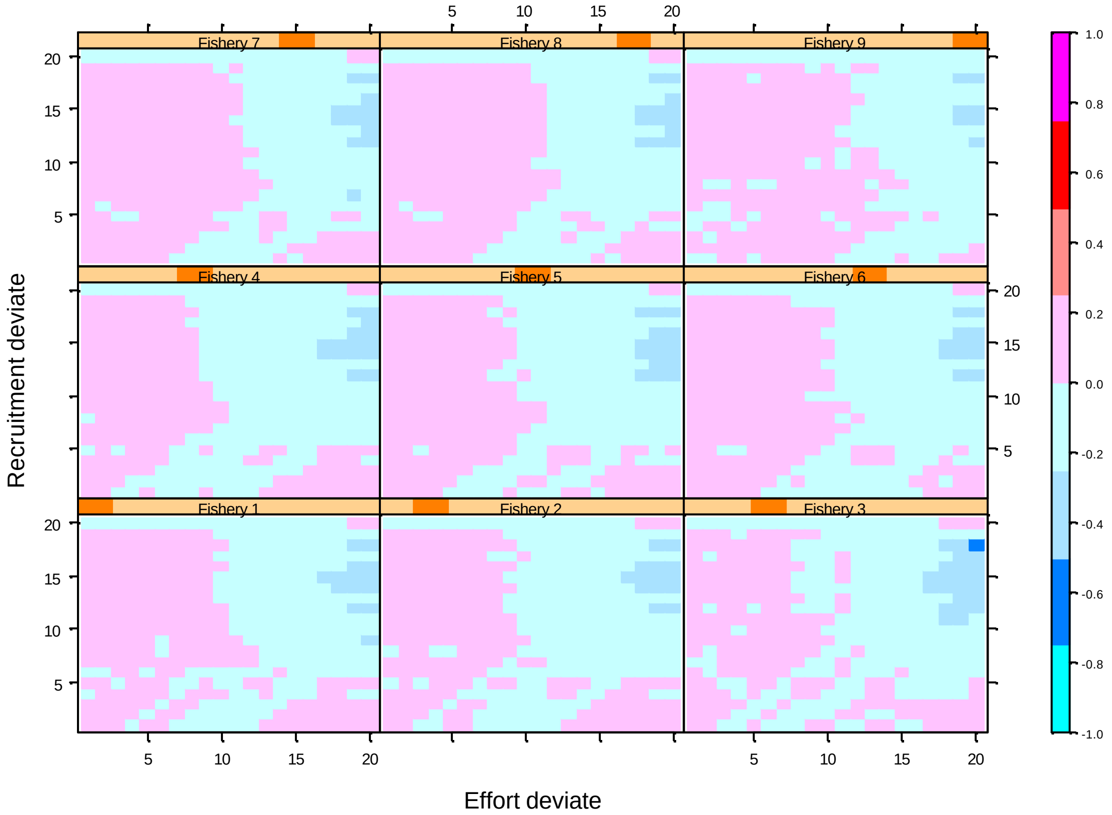


Figure 3. Correlation between the estimate yellowfin tuna recruitment and estimated effort deviates for the surface fisheries.

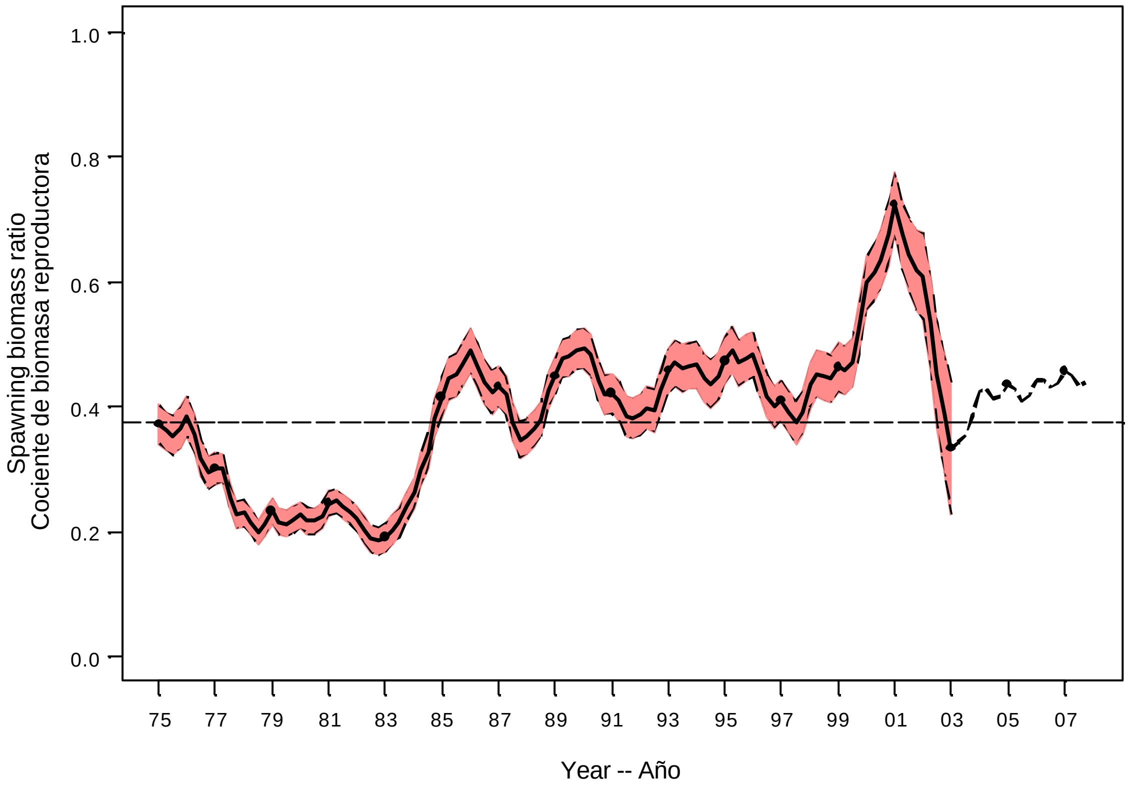


Figure 4. Estimated time series of spawning biomass ratios (SBRs) for yellowfin tuna in the EPO. The dashed extension to the solid line represents the projected SBR under current effort and average recruitment. The thin lines represent approximate 95% confidence intervals. The dashed horizontal line (at about 0.37) identifies the SBR at AMSY.

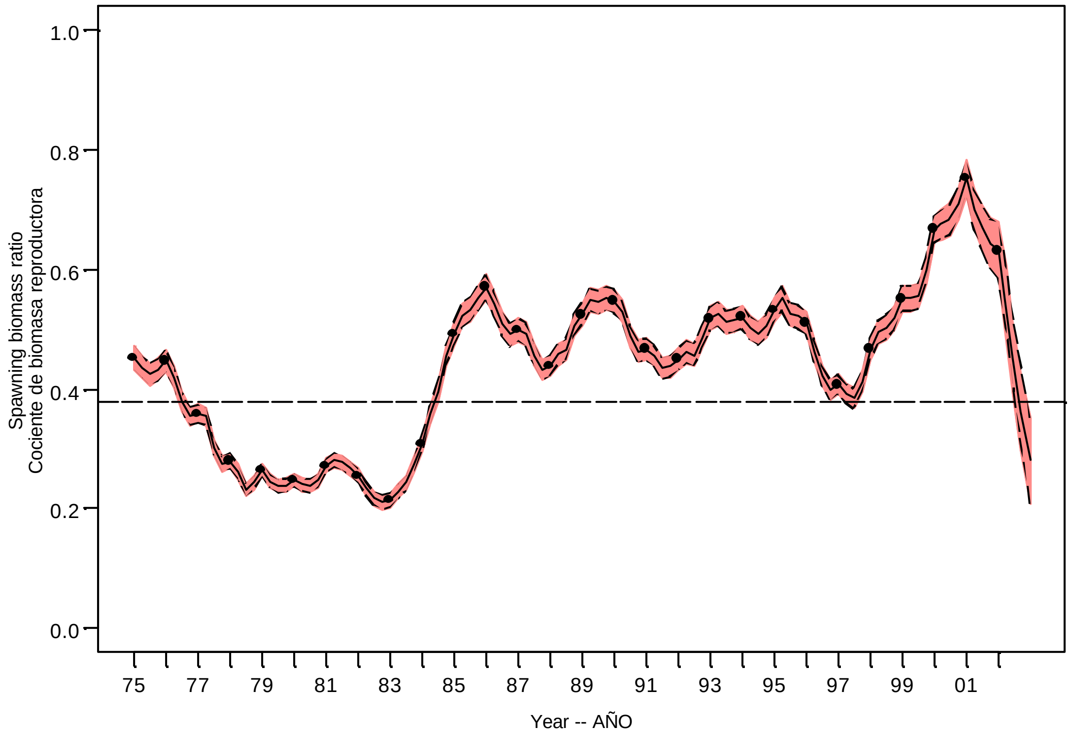


Figure 5. SBR and the associated confidence intervals for the iterative reweighting sensitivity.

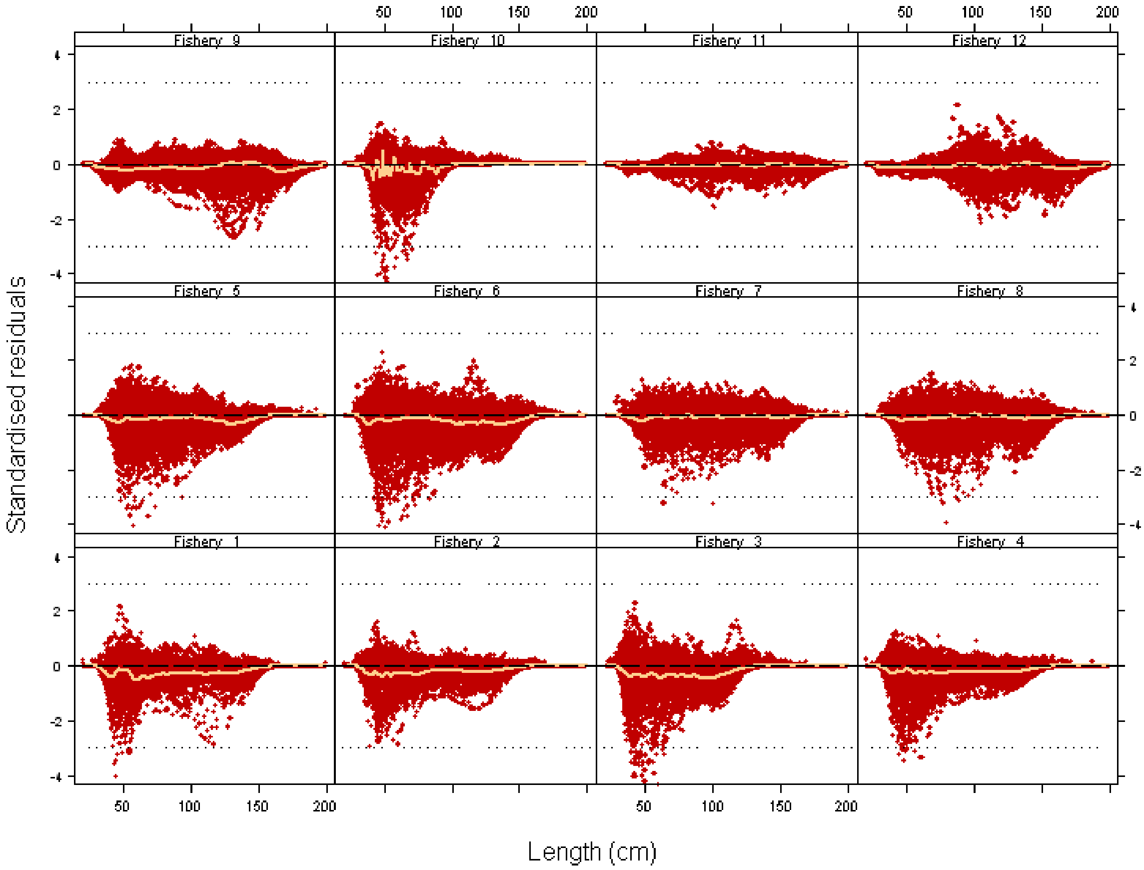


Figure 6. Deviates of the fit to the length-frequency data when using the basecase sample sizes for the yellowfin tuna assessment.

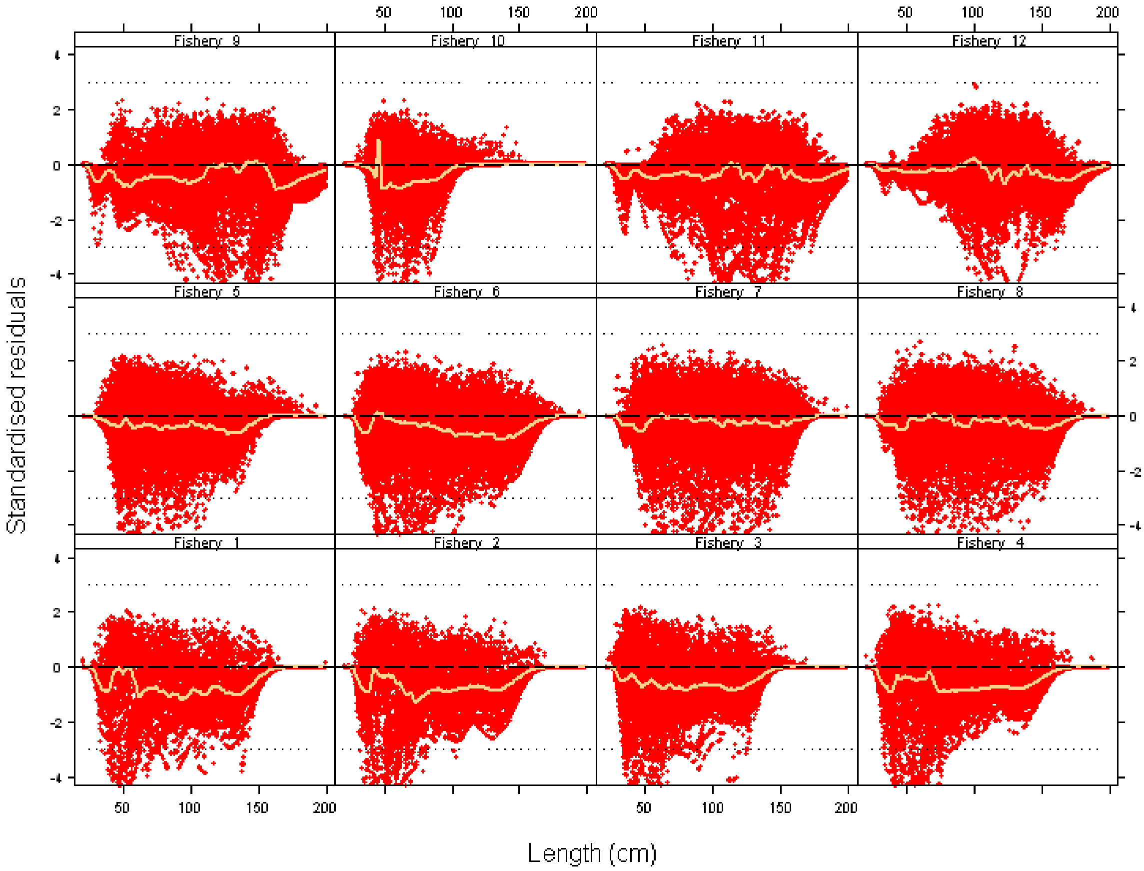


Figure 7. Deviates of the fit to the length-frequency data when using the basecase sample sizes for the yellowfin tuna assessment.

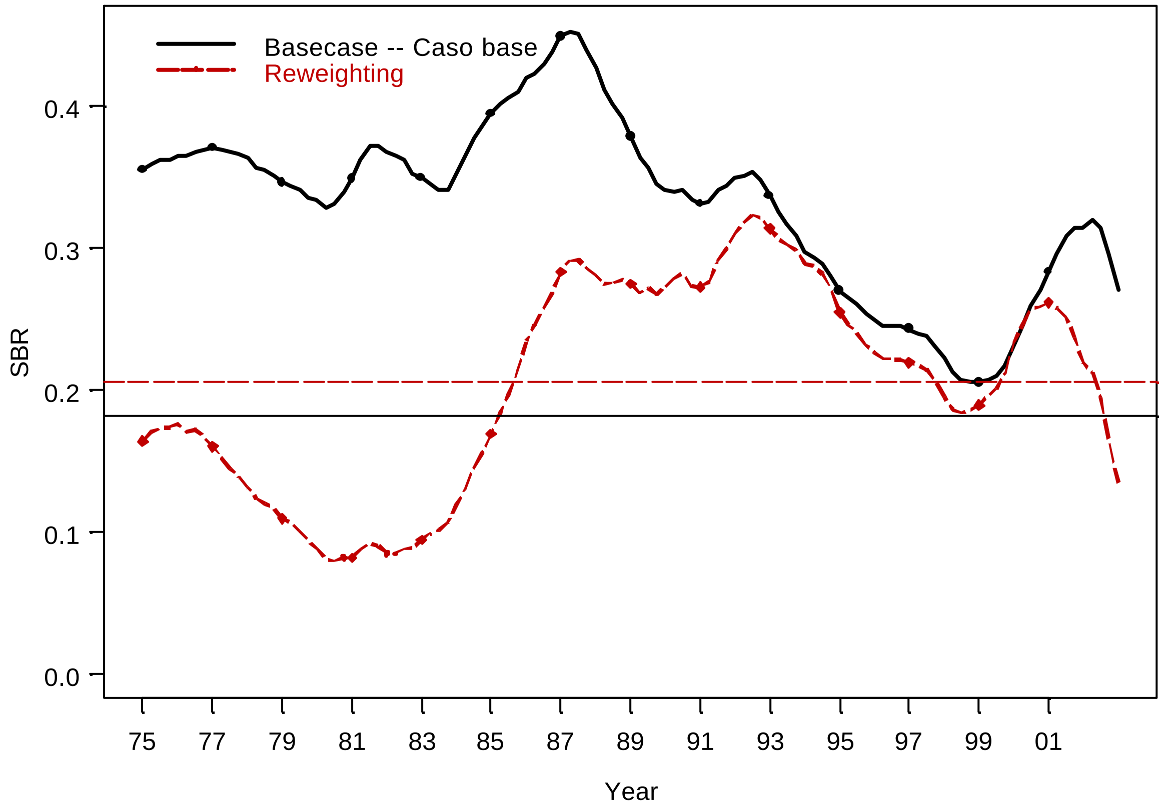


Figure 8. Comparison of estimates for bigeye tuna of the spawning biomass ratio (SBR) from the base case and with the length frequency sample sizes based on the iterative re-weighting procedure. The horizontal lines represent the SBR associated with AMSY.

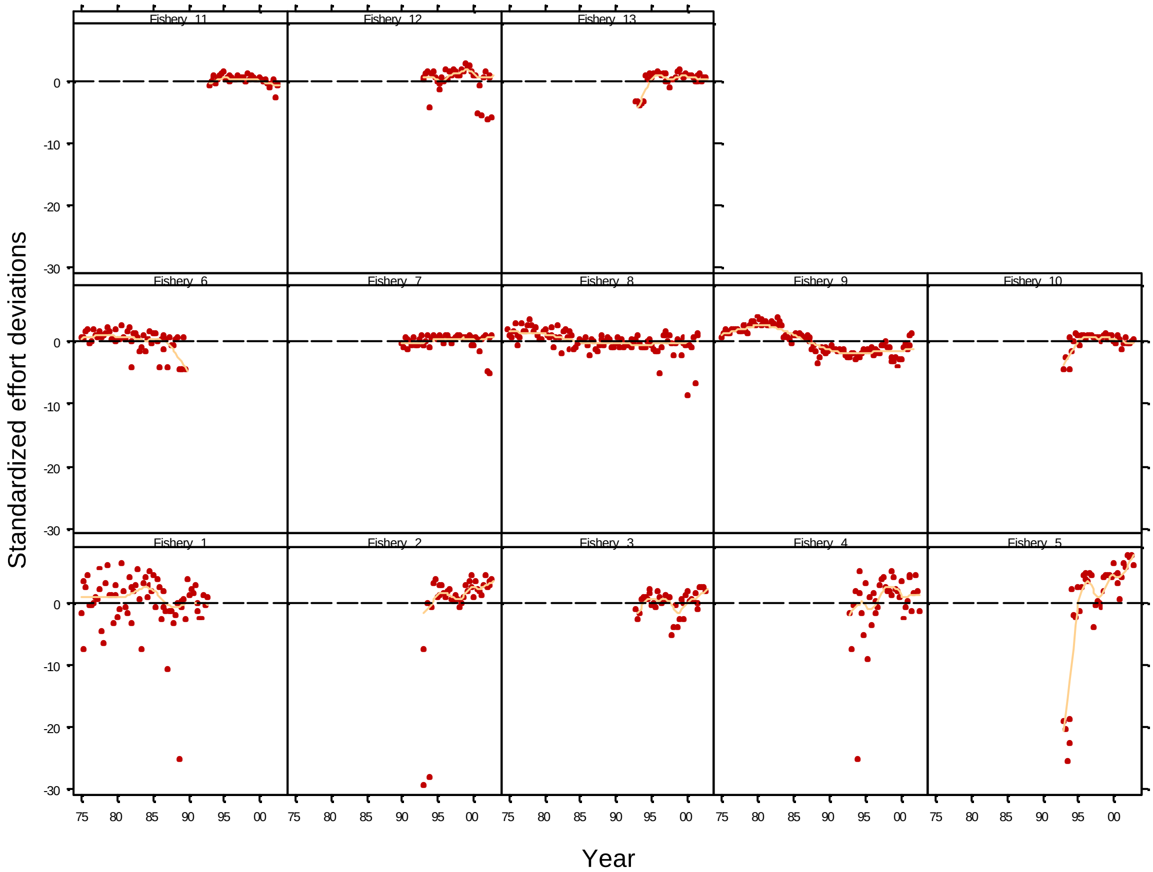


Figure 9. Standardized effort deviates for bigeye tuna by fishery and time quarter when the length frequency sample sizes are based on the iterative re-weighting procedure. The fitted line is a loess smoother. Fishery 9 is the southern longline fishery.

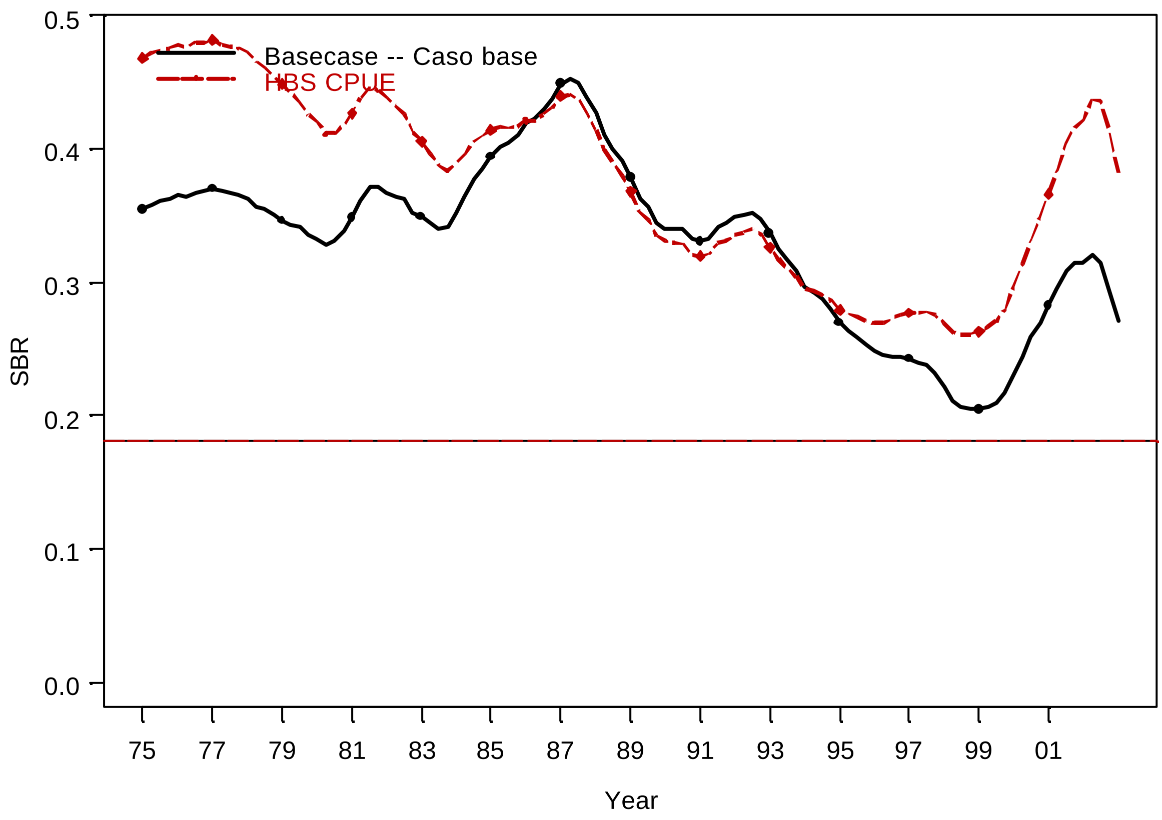


Figure 10. Comparison of estimates for bigeye tuna of the spawning biomass ratio (SBR) from the base case and the sensitivity with the habitat standardized CPUE as used in the last assessment (Maunder and Harley 2002). The horizontal lines represent the SBR associated with AMSY.

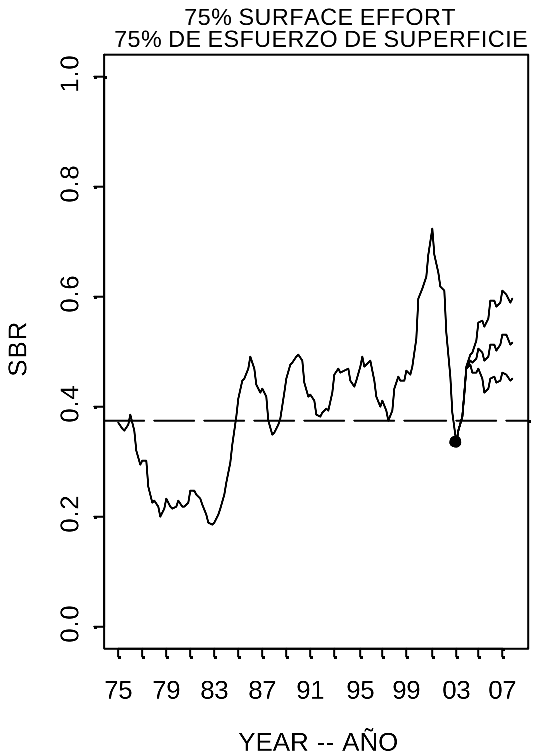


Figure 11. Projections for yellowfin tuna using the old method that only includes uncertainty in future recruitment.

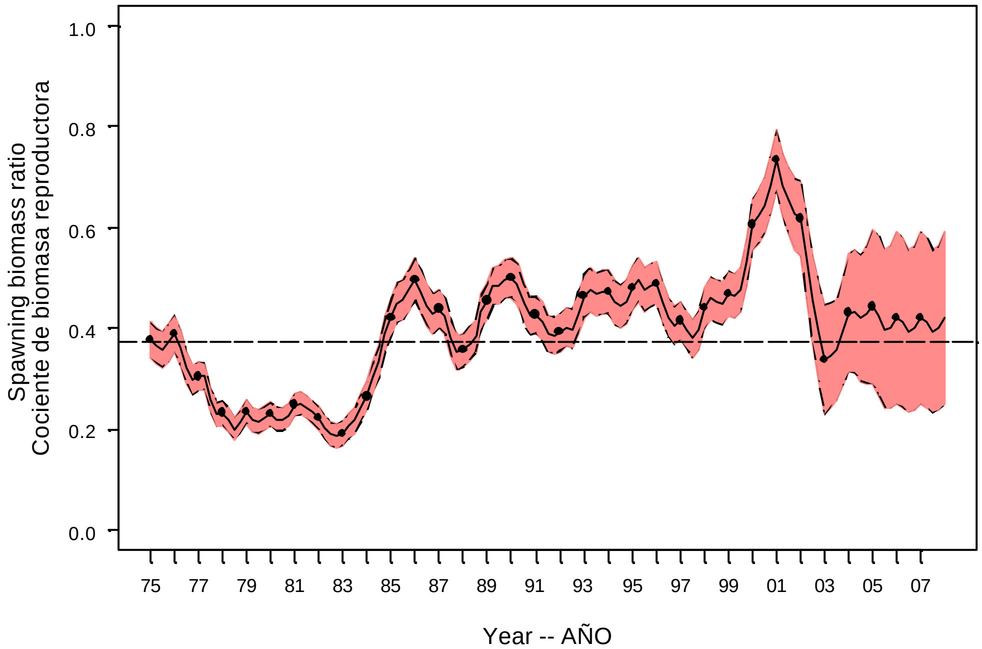


Figure 12. Projections for yellowfin tuna using the normal approximation to the likelihood profile method that includes uncertainty in future recruitment and parameter uncertainty.