Review of National Marine Fisheries Service Framework to Estimate Growth Rates of Eastern Tropical Pacific Dolphins Caught in Tuna Purse-seine Fisheries

by

Dr. Murdoch McAllister
London, United Kingdom

Review organized by the Center for Independent Experts, University of Miami

May 31, 2002

Executive Summary

1. The overall approach to estimating growth rates, abundance and depletion of eastern tropical Pacific dolphins (ETPD), the northeastern spotted (*Stenella attenuata*) and eastern spinner dolphins (*S. longirostris*), is adequate and appropriate. Alternative population dynamics models are fitted to absolute indices of abundance obtained from research vessel surveys and line transect methods carried out since 1979 and a Bayesian statistical methodology is appropriately applied to probabilistically estimate population parameters. This statistical modeling approach is arguably the most rigorous and conceptually consistent of the ones available for this problem.

2. The general types of population dynamics models considered, the simple exponential, generalized logistic and particularly the Leslie Matrix age-structured model, are appropriate for modeling trends in abundance and growth rates of ETP dolphins. However, modifications to them are required to more adequately evaluate the factors impeding population recovery.

3. A major weakness in the assessment is the large gap from 1991-1997 in the line transect abundance time series. This leaves only three data points in 1998-2000 for statistical inference over the last decade. Without using other data that span this gap, evaluation of the abundance trend in it, and possible factors impeding population recovery will have a weak empirical basis.

4. It is indisputable that there is trend-bias in the fishery dependent (fd) indices of abundance (i.e., a net annual deviation in the annual rate of change in the fd time series from the annual rate of change in the population). However, statistically rigorous and scientifically sound approaches exist and should be applied to remove the trend-bias in these time series so that they can be used in the stock assessment and help reduce uncertainties over the recent apparent lack of recovery.

Recommendations to be considered for implementation immediately

1. It is recognized that this first recommendation is controversial, and deviates from my promise at the review meeting that there will be no new surprises in my recommendations. However, as a result of further evaluation of the issue, it is recommended that the NMFS considers including in the current stock assessment approach for each species both the fishery independent (fi) and ***bias-corrected*** fishery dependent (fd) indices of abundance (rather than only the fi indices at present). It is recommended that within the stock assessment model, the annual rate of change of bias in the fd indices of abundance be estimated using a linear model for trend-bias while assuming no trend-bias in fi indices. This should help to improve the empirical basis for the current estimates of growth rate and the relatively poor ability to discriminate among alternative models because of large gaps in the data, especially the lack of abundance data from 1991-1997. If a rigorous approach to removing the trend-bias from the fd indices is applied, the issue of assigning weights to the fi and fd becomes less contentious and empirically estimated variances based on fits of the models to the data could be applied. However, this should be done applying the constraint that the fd indices are given no more weight in the estimation than the fi indices (i.e., the stock assessment model variance for the fd data should be constrained to be no less than that for the fi data). It is also recommended that NMFS review this issue together with IATTC to come up with a mutually accepted approach to dealing with the fishery dependent data in the ETPD stock assessment.

2. It is recommended that an additional age-structured model be developed that models incidental mortality rates on one-year-old calves as a function of the annual per capita index of exposure to dolphin purse-seine sets, the latter measured by the number of dolphin sets on the species x the annual average number of animals caught or chased per set divided by the population abundance.

3. When different population dynamics model structures were compared, the criterion for choice was AIC. Although AIC is widely accepted and considered a rigorous and objective criterion, it is difficult to interpret and leads to only one model being selected. It is recommended that Bayes' marginal posterior probabilities (i.e., Bayes' factors) be computed instead for each alternative model considered. These latter

statistics are more rigorous and useful because unlike AIC they account for parameter uncertainty and have a probabilistic interpretation, e.g., the probability that each model is true when compared to the others, given the data. And unlike AIC, these probabilities can and should be applied in a decision analysis of alternative policy options.

4. It is recommended that a scenario-based approach be applied to evaluate the plausibility of various factors that might have been impeding population recovery over the last few decades since the reported kills in dolphin sets were substantially decreased. Bayes' marginal posterior probabilities should be computed and presented with each alternative population dynamics model to indicate the relative credibility/plausibility of each given both the fi and fd data.

5. It is recommended that particular attention be given to evaluating the plausibility of the fishery-induced calf-mortality model (Recommendation 2) against that of other models that do not implicate the tuna-purse-seine fishery as the chief cause for impeding ETPD recovery.

Recommendations to be considered in the near future

1. It is recommended that the estimation performance of alternative estimators be evaluated by repeatedly simulating data with known model structures and values for model parameters, applying each estimator to estimate the parameters and then computing the bias and precision in each alternative estimator. This should be applied to evaluate the relative improvement in estimation performance of estimators that use both the fi and fd indices of abundance versus only the fi indices of abundance.

2. It is recommended that the fishery independent indices of abundance be treated as relative instead of absolute indices of abundance and that an informative prior probability distribution be constructed for the constant of proportionality that relates the true abundance to the relative abundance indices.

3. It is recommended that the "mu-model" which is a perfectly sensible and useful model to apply, be renamed and not discarded but still applied as an explanatory and predictive model for ETPD population dynamics, along with the various other models recommended and already applied.

4. It is recommended that a formal Chi-square statistical measure of model deviance be computed for each estimation as a diagnostic of the goodness of fit of the model to the data.

5. It is recommended that both the spotted and spinner dolphin stocks be modeled simultaneously as separate stocks in the same population dynamics model to estimate parameters that could be considered to be similar or the same between the two populations. One such parameter could include a parameter expressing a decline in reporting rates of dolphins killed in purse-seines since 1992 (analogous to "mu" in the soon to be renamed "mu-model"). Other parameters that could be considered to be similar or the same could include Rmax, the breakpoint year in the breakpoint models, and the relative deviation in Rmax, and K after the breakpoint, in the breakpoint models. As mentioned above, Bayes' factors and marginal posteriors could be computed to compare the relative credibility of the alternative models considered, even when populations are modeled separately or within the same model. The likelihood function used should take into account the lack of independence of the annual abundance estimate for the two stocks because they are estimated from the exact same survey.

Background[1]

The tuna purse-seine fishery has used the association between tuna and dolphins to fish in the eastern tropical Pacific Ocean for over five decades. Three stocks of dolphins were depleted by high historical levels of dolphin mortality in tuna purse-seine nets, with an estimated 4.9 million dolphins killed during the fourteen year period 1959-1972. After passage of the Marine Mammal Protection Act (MMPA) in 1972 and the increased use of fishing equipment and procedures designed to prevent dolphin deaths, mortality decreased during the late 1970s, 1980s and 1990s to levels that are generally considered biologically insignificant.

While changes in the fishery have greatly reduced the observed mortality of dolphins dramatically, the MMPA, as amended by the International Dolphin Conservation Program Act, requires that the National Marine Fisheries Service (NMFS) conduct research consisting of three years of population abundance surveys and stress studies to form the basis of a determination by the Secretary of Commerce regarding whether the "intentional deployment on, or encirclement of, dolphins by purse-seine nets is having a significant adverse impact on any depleted dolphin stock". Specific to this review, NMFS must essentially determine whether or not the depleted dolphin stocks are recovering, and if so, at what rate and at what level of certainty.

The topic of this review is the overall framework that will be used to estimate the growth rate of two dolphin populations of interest, the northeastern offshore spotted dolphin and the eastern spinner dolphin. The framework uses growth rates estimated by fitting a population model to the three-year survey estimates and other available estimates of abundance. Estimates from research vessel surveys using line transect methods are available for three periods: 1979-83 (four estimates), 1986-90 (five estimates), and 1998-2000 (three estimates), for a total of twelve estimates over twenty-one years. Reviewers will also be asked to evaluate the inclusion or exclusion of a set of fishery-dependent indices of abundance, resulting from data collected by observers onboard tuna vessels. Two types of population growth rate will be estimated: (1) exponential rate of change from 1979-2000 and (2) intrinsic rate of increase under the assumption of a density-dependent model where pre-exploitation population size in 1958 is considered carrying-capacity. Both an aggregated population model and an age-structured model will be used. Bayesian statistics, using a numerical integration method, will be used to estimate a probability distribution for the population growth rate.

In this report I have evaluated the following.

(1) The overall framework that will be used to estimate the population growth rates of northeastern spotted dolphins and eastern spinner dolphins.

(2) The adequacy of the alternative population dynamics models used to model growth rates.

(3) The adequacy of model inputs, particularly, the exclusion or inclusion of a set of fishery-dependent indices of abundance, resulting from data collected by observers onboard tuna vessels.

(4) The adequacy of the statistical methods, particularly the Bayesian methods, used to estimate growth rates.

(5) The adequacy of interpretations of the stock assessment results and protocols used for model selection when alternative models with competing explanations for historic population dynamics are fitted to the same data.

[1] The first three paragraphs were copied directly from the statement of work, which is attached in Appendix 2.

(6) The adequacy of procedures and diagnostics used to evaluate the estimation performance of the estimators applied, the goodness of fit of models to the data and the computational performance of algorithms to compute Bayesian posterior distributions.

Although the methods adopted for the estimation of population growth rates are generally adequate and appropriate I have provided recommendations to help improve the estimation of growth rates and account for the recent trends in abundance series available. As seen in the executive summary, the recommendations are prioritized in terms of changes that are recommended to be considered for implementation immediately and other changes that could be implemented in the near future.

Description of Review Activities

The review activities consisted of four different phases. The first was a review of documents provided. The second consisted of taking part in meetings with National Marine Fisheries Service (NMFS) scientists working on ETP dolphins at the Southwest Fisheries Science Lab in La Jolla on April 3 and 4, 2002. The third consisted of further requests by e-mail to NMFS scientists for further information. The fourth consisted of writing the review report.

1. Review of papers

To obtain a background to the stock assessment, several papers were provided on the website, http://swfsc.nmfs.noaa.gov/prd/assessment/docs.html. These papers and others provided at the La Jolla meeting are listed in Appendix 1. The paper:

Wade, P. 2002. Assessment of the population dynamics of the northeast offshore spotted and the eastern spinner dolphin through 2001. Doc. prepared for the Southwest Fisheries Science Center, National Marine Fisheries Service. La Jolla California.

provides an outline of the methodology to be used for estimating growth rates, abundance and depletion of ETPD populations and some preliminary results. This is the main work to be reviewed. The review of this work is described in detail below.

2. Meeting with National Marine Fisheries Service Scientists in La Jolla, April 3 and 4, 2002

The second review activity consisted of meetings between the two reviewers, Dr McAllister and Dr Malcolm Haddon, and National Marine Fisheries Service (NMFS) scientists working on ETP dolphins at the Southwest Fisheries Science Lab in La Jolla on April 3 and 4, 2002. Those also attending included the following:

Dr. Paul Wade - presenter of assessment model
Dr. Steve Reilly - IDCPA research coordinator
Dr. Tim Gerrodette - generated the dolphin abundance estimates
Dr. Bill Perrin - ETP dolphin specialist and pioneer of the dolphin-tuna purse-seine problem
Dr. Wayne Perryman - dolphin photogrammetry specialist
Dr. Lisa Ballance - ecosystem specialist
Dr. Dave Bratten – IATTC
Dr. Bill Fox, chief scientist of the Washington branch of NMFS
Dr. Eric Archer - biologist
Dr. Michael Tillman, Director of SWFSC
Mr. Paul Fiedler - oceanographer
Ms. Nicole le Boeuf, fisheries-science-policy aide, NMFS Washington
Ms. Meghan Donahue, biologist, SWFSC

It should be noted, however, that this list is not fully inclusive and my apologies to those whose names I have left out due to failing to catch their names while in attendance.

In this meeting presentations were made by NMFS scientists to give a background to the issue, present the modeling work carried out, present methods used to develop inputs for the modeling, and outline evidence about various phenomena that could potentially explain the recent trends in abundance. An open and flexible format was followed in which the reviewers could ask questions, make comments and recommendations, and make requests to see various model outputs, inputs, additional papers, and other datasets that were deemed to be relevant to the review. The main events of the meeting were as follows:

(1) Morning April 3. Presentation by Dr. Steve Reilly: Background to the tuna-dolphin problem.

(2) Afternoon April 3. Presentation by Dr. Paul Wade: Bayesian statistical methodology used to estimate growth rates of northeastern spotted and eastern spinner dolphins.

(3) Morning April 4. Presentation by Dr. Paul Wade: Modeling results obtained on age-specific selectivities of northeastern spotted dolphins and discussion of using simple parametric models for selectivity. Also discussion of the mathematical form of the generalized surplus production model used to estimate recent population growth rates of ETP dolphins.

(4) Morning April 4. Presentation by Dr. Tim Gerrodette: Review of line transect and distance estimation methods used to estimate northeastern spotted and eastern spinner dolphin abundance.

(5) Morning April 4. Presentation by Dr. Paul Wade: Review of methods used to estimate bycatch of dolphins in the Eastern Tropical Pacific Purse-seine fisheries for tuna 1959-1972.

(6) Morning April 4. Presentation by Dr. Steve Reilly: Outline of recent findings from studies on the stress imposed by chase and capture of dolphins by tuna purse-seine vessels.

(7) Morning April 4. Presentation by Dr. Steve Reilly: Outline of recent findings from ecological studies on ETP dolphins.

(8) Afternoon April 4. Viewing of video with over-view by Dr. Dave Bratten: Documentary on the capture of ETP dolphins in tuna purse-seine vessels.

(9) Afternoon April 4. Discussion of indirect effects of tuna purse seining on mortality rates of ETP dolphins, particularly the calves and the potential use of the per capita number of dolphin sets per year as a covariate for dolphin calf mortality rates.

(10) Afternoon April 4. Discussion of the issue use of whether the fishery dependent indices of abundance for spotted and spinner dolphins should be used in assessing their growth rates and how these particular indices of abundance should be included in the stock assessment modeling.

The information provided by e-mail is included in Appendix 3.

Summary of Findings

Overall, the general framework applied to determine whether or not the depleted stocks of ETP dolphins are recovering, at what rate and at what level of uncertainty is adequate and appropriate. Several different recommendations have been developed to improve the use of available data and other information. These

recommendations are prioritized in terms of ones that should be considered for implementation immediately and others that could be implemented in the near future.

(1) The overall framework that will be used to estimate the growth rate of northeastern spotted dolphins and eastern spinner dolphins.

The framework to estimate population growth rates of the eastern spinner and northeastern spotted dolphin populations applies Bayesian estimation and fits a population dynamics model to fishery independent indices of abundance.

Is the Bayesian stock assessment approach appropriate for this problem?

The Bayesian statistical approach using the SIR integration method (Rubin 1988) is applied when fitting the population dynamics models for ETPDs to abundance data to determine whether or not the depleted stocks of ETP dolphins are recovering, at what rate and at what level of uncertainty. The Bayesian approach to estimation allows uncertainty in model parameters such as growth rate and variables such as abundance to be quantified in the form of probability distributions. In contrast to the alternative Frequentist approach, it formally allows the value of a parameter to be treated as a random variable (RV). Because of this probability distributions can be defined for model parameters and quantities derived from them (e.g., population abundance). The Bayesian statistical approach is empirically rigorous because it requires that initial (or prior) probability distributions for model parameters be updated formally with newly obtained data to obtain posterior probability distributions. The approach permits the inclusion of a wide variety of types of data and information about the population to be rigorously incorporated in the form of probability distributions. This estimation approach has recently become a mainstream approach to estimation in the field of statistics and has become widely accepted as a rigorous estimation and decision analysis methodology in fisheries stock assessment (Punt and Hilborn 1997; McAllister and Kirkwood 1998). Wade (1999) provides convincing analyses in support of the Bayesian approach for abundance estimation and the approach is arguably the most suitable approach of all to deal with this particular stock assessment problem. For the above reasons, it is far more appropriate than standard Frequentist approaches that apply maximum likelihood estimation and bootstrapping. Although the Bayesian methods applied are appropriate and statistically rigorous they could be enhanced with some refinements that are described below.

Should line transect abundance estimates be used as absolute abundance estimates?

The line transect abundance estimates used for the estimation are treated as absolute estimates of abundance. Given the well-conducted line transect surveys and rigorous distance estimation methodologies applied (Gerrodette and Faucada 2002), the use of the estimates as absolute indices of abundance in the stock assessment is reasonable. However, unless the assumption is thoroughly evaluated, marine scientists can never be certain, even with the most rigorous survey and estimation methods, that the values that they calculate are unbiased estimates of abundance. This is especially so for wide-ranging animals such as dolphins that have complex behaviours and occur in groups of widely varying sizes. For example, from helicopter observation, it is known that each observer produces consistently biased estimates of group size. Helicopters were present to evaluate the observer bias in the latter years of the survey but missing in the earlier years of the survey. Bias-correction values for group size estimates based on helicopter-evaluated observers have been extrapolated to observers for whom such external validation was not done. This makes arguments about the unbiasedness of values in the earlier part of the series a little less tenable. Additionally, a parametric density function is applied to describe the proportion of sightings at each perpendicular distance from the line transect. Imperfections in the choice of the density function, especially if it does not fit very well the observations close to the trackline, could also result in biased estimates of abundance.

Although scientists appear to be convinced that the line transect estimates for ETPD have little or no bias, it is still appropriate to develop an alternative stock assessment in which they were treated as relative abundance indices. The Bayesian approach to estimation that is applied would allow for informative prior density functions to be developed for the constant of proportionality between population abundance and the line transect indices of abundance. While this approach has been applied to acoustic (McAllister et al. 1994; Boyer and Hampton 2001), trawl survey (McAllister and Ianelli 1997), commercial trawl (Branch 2001) and pelagic egg estimates of abundance (EC project, QLRT-PL1999-01253), it has not yet (to my knowledge) been applied to line transect survey estimates where sightings and distance estimation methods are applied. Thus it is recommended that attention is given to developing an informative prior density function for the constant of proportionality between actual abundance and the line transect index of abundance and treating the line transect estimates as relative indices in the stock assessment.

(2) The adequacy of the alternative population dynamics models used to model growth rates.

Three different types of population dynamics models were used to estimate population growth rates. These include a simple exponential growth rate model:

$$N_t = N_{t-1}\exp(r) - H_{t-1}$$

a generalized surplus production model (GSPM):

$$N_t = N_{t-1}\left(r_{\max}\left(1-\left(\frac{N_{t-1}}{K}\right)^Z\right)\right) - H_{t-1}$$

and a simple age-structured model (not described mathematically here). N_t denotes the abundance in numbers in year t, r denotes the rate of change, r_{max} denotes the maximum growth rate, K denotes the carrying capacity, H_t denotes the number of animals killed in the fishery in year t, and z determines the fraction of K at which surplus production is maximized.

The use of three different population dynamics models is sensible because it allows the evaluation of the sensitivity of estimates of growth rate to the structural model form adopted. Providing that reliable age-structured information are available, e.g., on rates of natural mortality, fecundity, density-dependence, and the relative vulnerability of different ages to fishing gear, the age-structured model offers the most rigorous modeling approach to evaluate hypothesized population processes.

<u>The form of the Generalized Surplus Production Function (GSPF)</u>

There are several alternative formulations of the GSPF, and the mathematical behaviours of each are not identical. The alternatives include the Fletcher model (Quinn and Deriso 1999), and the Gilbert (1992) version of the Pella Tomlinson model. The form used in this analysis has the property that for a given value for K, the maximum sustainable yield (MSY) increases progressively as the inflection point in the surplus production function increases from 0 to carrying capacity (K). In contrast, under the Fletcher model, MSY is independent of the inflection point. In the La Jolla meeting, the suitability of the formulation used for ETP dolphins was questioned. Paul Wade pointed out that theoretical modeling research (Taylor and deMaster 1993) supported the notion that MSY should increase as the inflection point approached K. While this general form might be sensible, the model prescribes for a given intrinsic rate of increase (r), K and z (parameter that determines the inflection point), a particular value for MSY. It remains uncertain whether this particular functional form conforms closely to the actual surplus production function for ETP dolphins. It is recommended that the posterior median surplus production

function obtained from the age structured model be compared with the posterior median surplus production function obtained from the generalized surplus production function to see whether they are reasonably similar. If they are not, it might be appropriate to consider some other functional forms for the generalized surplus production that could more closely mimic the ones given under the age structured model.

Age-structured model

The age-structured model used was in the form of a density-dependent Leslie matrix. Density dependence was modeled in fecundity using a logistic formulation:

$$f_t = f_0\left(f_m - f_0\right)\left(1-\left(\frac{N_t}{N_{eq}}\right)\right).$$

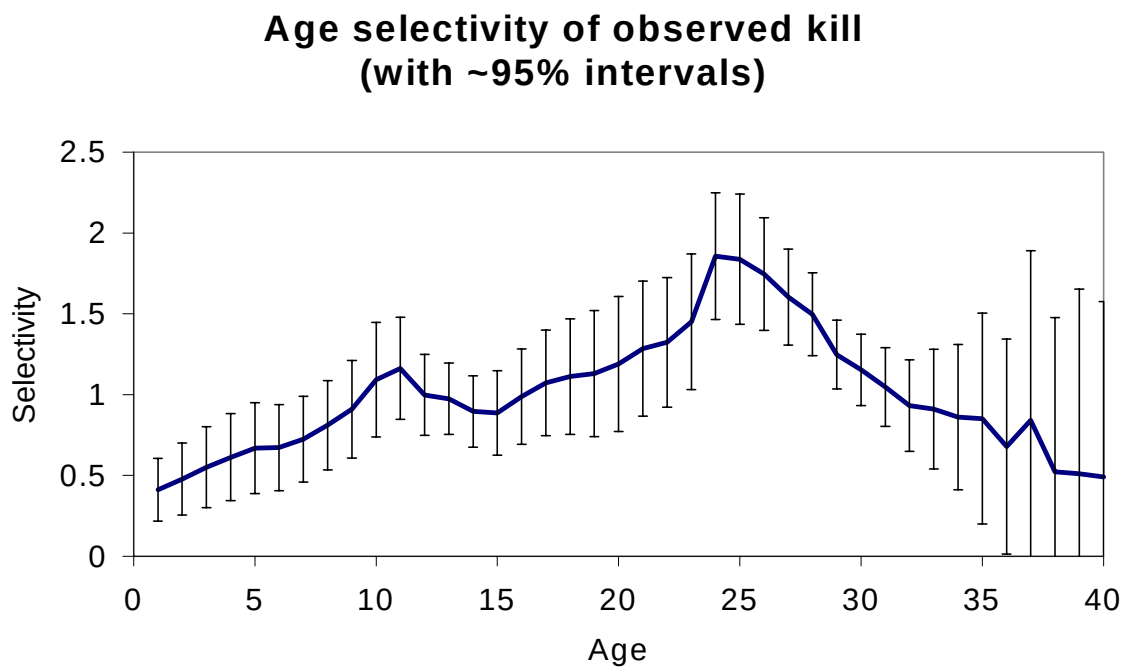
This is a relatively simple formulation that should work adequately in deterministic population dynamics models. However, if a stochastic or state space (Schnute and Richards 1995) model were to be applied (a further recommended improvement), then this model might give the anomalous result of negative fecundity if N_t happened at some point to exceed N_{eq}.

The vulnerability at age was modeled to be a unique fitted parameter for each age class. The set of values chosen was found by finding the best fitting values once the other model parameters were chosen. The data used are observations of the number of animals observed to be killed in each age group from samples of animals killed 1976-1982 (Fig. 1a).

Fig. 1. a) Age distribution of kill of NEOSD from years 1976-1982. b) Estimated vulnerability at age. Source: P. Wade.

a.

b.



The vulnerability at age estimation approach could be improved. First uncertainty in the vulnerability at age is not very adequately taken into account because only the best fitting values for vulnerability are chosen given each particular set of the values for the other model parameters. Second, because there are no constraints on the potential values for relative vulnerability at age and a separate value is obtained for each age, the model may be overfitting the data and give unrealistic estimates of vulnerability at age with some sets of values for the other model parameters.

Purse-seine fishing practices have changed considerably since 1976-1982 and it is questionable whether the same vulnerability at age pattern that might have occurred back then continues to be the same in recent years. For example, after the implementation of the International Dolphin Conservation Program in 1992, all participants agreed to have observers placed on vessels. Techniques to reduce the number of animals killed have also been improved. The issue of potential temporal changes in selectivity should be formally addressed by NMFS biologists. If there are good reasons to believe that the vulnerability at age has changed, then some plausible modeling scenarios need to be constructed and evaluated to take this into account and evaluate the potential impacts on estimates of population growth rates. There are comments on this issue further below.

It was suggested at the La Jolla meeting that some parametric function for vulnerability at age be developed and applied instead of the non-parametric function. This would involve the adoption of some particular functional form for selectivity and require the estimation of its parameters in the same manner to the estimation of the values of other model parameters. At the meeting, a posterior estimate of the vulnerability (or selectivity) at age was provided. This curiously had two humps, one at about 12 years and the other at about 25 years and a pronounced decay after the age of 25 years (Fig. 1b). The age of maturity is between 11 and 12 years and there could be some linkage in this and the observed hump in selectivity at these ages. The data on the age distribution of the kill becomes quite sparse after the age of about 25 years. Also, the approximate 95% probability intervals on the selectivity become very wide after about 30 years. This decline after age 25 years could potentially be an artifact of model structure. The rate of natural mortality is estimated for adults and assumed to be constant with age. The estimated

decline in vulnerability after the age of 25 could be spurious because of the sparseness of the data and estimates of vulnerability of older animals being confounded with a possible effect of senescence. The plausibility of declining vulnerability at age after the age of about 25 years relative to the possibility of an increase in the rate of natural mortality should be evaluated by expert judgment since no data (to my knowledge) exist to sort this out. If both alternatives are plausible then some additional modeling scenarios should be constructed to evaluate the implications of both possibilities.

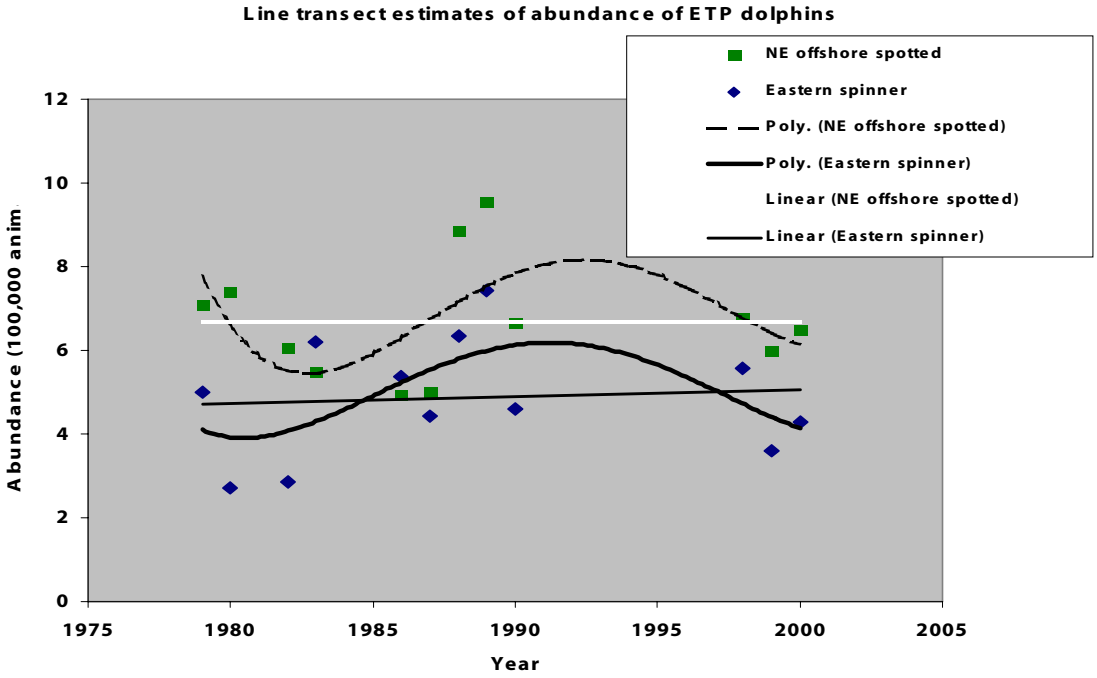
At the La Jolla meeting, a linear functional form for vulnerability at age was suggested. This would involve a fitting a line to the hump at about 12 years (a slope (positive) and y-intercept (positive)), a linear decline from this hump to 15 years (a slope parameter (negative)), and then a linear increase to the age of 25 years (a slope (positive)). Under the assumption of no decrease in vulnerability with age after 25 years, a level line would be assumed from 25 years on. This selectivity function would thus require four estimated parameters. As mentioned above, one scenario could instead involve fitting a line with a line with a negative slope starting at 25 years to allow for the possibility of decreasing vulnerability after 25 years (a slope (negative)). The signs are given to indicate the domain over which the priors for these parameters should be defined.

(3) The adequacy of model inputs, particularly, the exclusion or inclusion of a set of fishery-dependent indices of abundance, resulting from data collected by observers onboard tuna vessels.

Adequacy of Abundance Indices to which the population dynamics models are fitted

The main sets of data to which the model is fitted are time series of line transect data from research vessel cruises in two and three year sequences starting in 1979 (Fig. 2).

Figure 2. Line transect estimates of abundance of NEOSD and ESD together with linear and 4th order polynomial fitted trend-lines.



The largest and most problematic gap in these data occurs from 1991-1997 when no research cruises were carried out. Then in 1998, 1999 and 2000 three more cruises were carried out. The sampling CVs in the

abundance observations range from about 13% to over 40% (Gerodette and Forcada 2002). Standard distance sampling methodology (Buckland et al. 1993) has been applied to develop the abundance estimations from the vessel line transect sighting observations. To analyze the temporal patterns in these data, separate linear and polynomial functions were fitted to both time series (Fig. 2). From 1979-2000, there does not appear to be any distinct net trend in the observations for both the NE offshore spotted dolphin and the Eastern spinner dolphin. For both populations, there is a slight dip in the early 1980s, a slight increase at the end of the 1980s and then a decrease to the lower observations at the end of the 1990s (Fig. 2). The similarity in both trends is quite remarkable and unlikely to be mere coincidence. It could reflect analogous patterns in the population trends and/or covariance in research survey observation error.

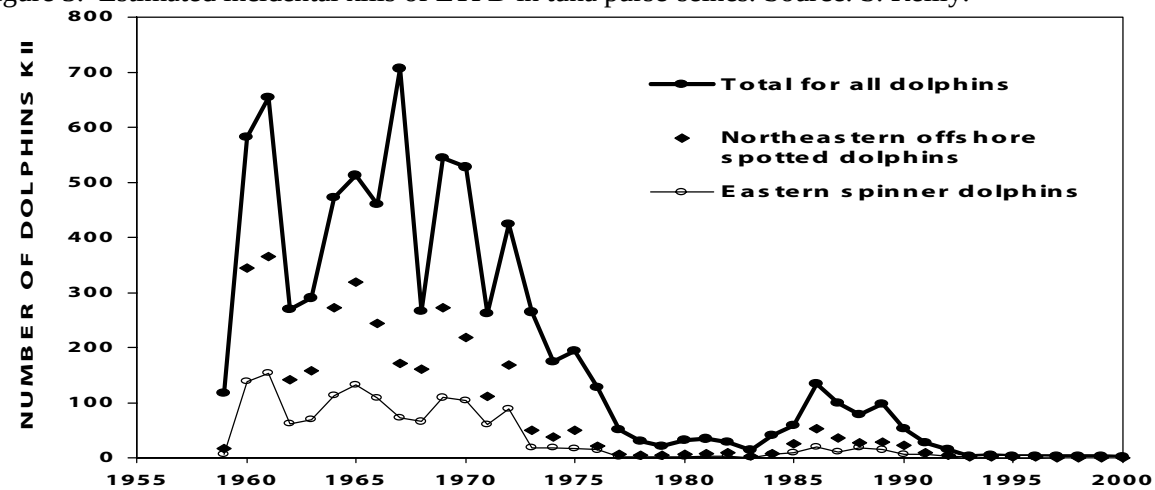
The fit of the assessment models to these data (Wade 2002) appears to be reasonable, except that there is some autocorrelation in the residuals. For example, for Eastern Spinner dolphin (ESD), the models over-predict in the late 1970's and under predict in the late 1980s. The abundance models, thus do not capture well the apparent rise in the late 1980s and then fall after that point. Due to the large gap in the line transect indices from 1991-1997, and without additional abundance data that could fill this gap, there is little empirical basis to evaluate the pattern in abundance for the 1990's.

The prior density functions used are in the form of uniform density functions with fairly wide ranges covering the ranges of values believed to be most plausible. As long as the actual values fall within the intervals provided and not on the edges of the intervals, this functional form for the priors is reasonable. It should be checked that the marginal posterior distribution, if updated from the prior does not give heaviest weights to values on or beyond the boundaries of the uniform priors. However, in the model output shown in the presentations, this does not appear to be the case. It might be advisable to consider instead non-uniform parametric density functions for the priors that are defined over the full range of plausible values and give weight to the more plausible values.

Adequacy of estimates of dolphins killed in tuna purse-seines

A key input to the population dynamics models includes the estimated annual number of dolphins of the species of interest killed incidentally in each year of the tuna purse-seine fishery. Estimates were precise since the formalization of the tuna vessel observer program in 1972 ($0 < 0.27$, source: P. Wade). Estimates were less precise for the period for 1959-1972 (Wade 1995) because of the low amount of observer coverage. CVs in estimates for NEOSD and ESD of these earlier kills ranged between 0.13 and 0.53. However, a rigorous estimation methodology was applied to estimate the numbers of dolphins killed and the CVs in the estimates during this period and it does not appear that the methodology can be significantly improved upon. The uncertainty in the estimated kills was appropriately taken into account in the Bayesian estimations by assuming that the errors of the mortality estimates were lognormal and perfectly correlated because mortality rates per set were pooled across that period. Further comments on the reliability of this estimated time series of incidental kills are provided below.

Figure 3. Estimated incidental kills of ETPD in tuna purse-seines. Source. S. Reilly.



Question of whether to use fishery dependent indices of abundance

Up until a few years ago, fishery-dependent indices of abundance were used in the stock assessments of ETPD assuming that they are unbiased indices of abundance (SWFSC 1999). However, recent studies (e.g., Lennert-Cody et al. 2001) have concluded that it is ill-advised to use these in population growth models and they are no longer used. The paper by Lennert-Cody et al. (2001) reviews in detail the estimation of dolphin abundance from fisheries data for the eastern spinner and NE offshore spotted dolphin populations. Their analyses suggested trends in biases causing an annual average decrease in each index of 1-1.5% per year.

The causes of the biases were concluded to result from progressive changes in data quality and fishery introduced-biases. The patterns of search for tuna have changed gradually over the years with direct search from the vessel decreasing, helicopter search increasing from 1980 and then remaining constant after about 1985 and radar search increasing dramatically from the mid-1980s. There has been a significant increase the number of helicopter sightings leading to sets. The pattern of reporting of sightings from crew to observers has shifted. In recent years it is believed that fewer of the sighted dolphin pods are reported to observers and that there is a tendency for crew aboard helicopters to selectively report dolphin herds associated with tunas. This is partly because the helicopter search methods have increasingly focused on large dolphin pods since these are more often associated with larger schools of yellowfin tuna. This implies that smaller pods sited by the crew tend to be reported less in the more recent years. Due to the increase in use of helicopters and radar for search and the longer distances between the vessel and detected dolphins, post-detection search has also increased and the rate of sighting in these is lower than in pre-detection search.

The authors concluded that due to temporal changes in the patterns in search and reporting of sightings there has been an increasing trend in data loss. The authors believe that it is conceivable that trends in bias and population growth rates may be of similar magnitude. They conclude that "trends in the index (or lack thereof) may reflect the confounding of trends in biases and changes in population size. Thus, we believe the use of these indices as input to population dynamics models is ill-advised."

In summary the following observations can be made about the Lennert-Cody et al. (2001) paper:

(1) The paper provides a detailed and rigorous analysis of factors in the data procurement and data analysis procedures for fishery dependent abundance indices that could lead to biases in the abundance trends (or lack thereof) indicated by the indices.

(2) Temporal changes in data quality that could cause spurious trends in the indices are demonstrated, e.g., as result of potentially anomalous trends in the proportion of sightings on and very near to the track-line.

(3) Temporal changes in fishing practices and data reporting practices that could and likely do cause spurious trends in abundance indices are also demonstrated.

(4) The effects of the sources of trend-biases appear to act in the same direction, to cause the calculated indices to have a net average annual decrease over time.

(5) The authors believe that the net average annual bias is an average decrease in each index by about 1-1.5% per year.

Several general points can be made about the issues addressed by this paper.

(1) It is likely that the trend information suggested by the fishery dependent indices is biased.

(2) The estimates of the indices themselves are fairly precisely estimated, despite being biased, particularly in the latter part of the time series. So whatever they are tracking they are tracking it quite precisely and based on the data and the distance estimation methods used to generate the indices, the effect of sampling imprecision on the values obtained is relatively small. The earlier years show considerably more imprecision due to fewer observations being available per year.

(3) Unless a reliable and statistically rigorous approach can be identified to remove from the indices the bias in the trend information they should not be used in stock assessment models.

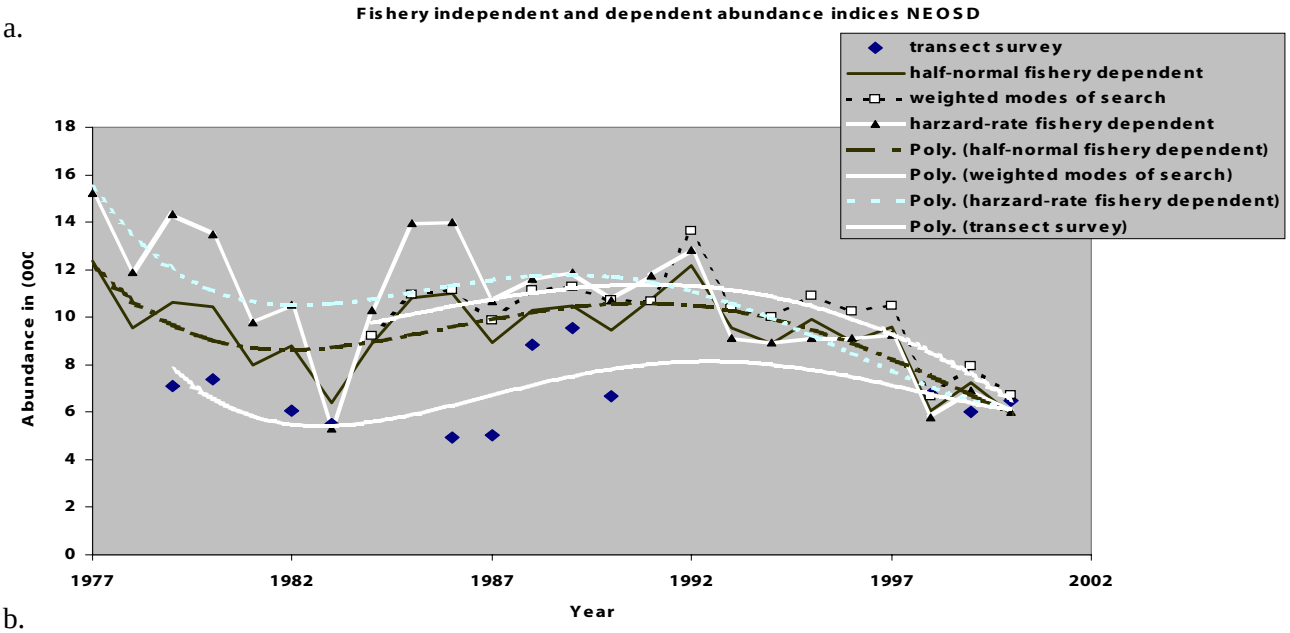
(4) Because the indices are computed from fisheries data and because of the non-random pattern of search of the fishing vessels, they should not be treated as absolute indices of abundance either.

Evaluation of temporal patterns in the fishery-dependent and fishery-independent data

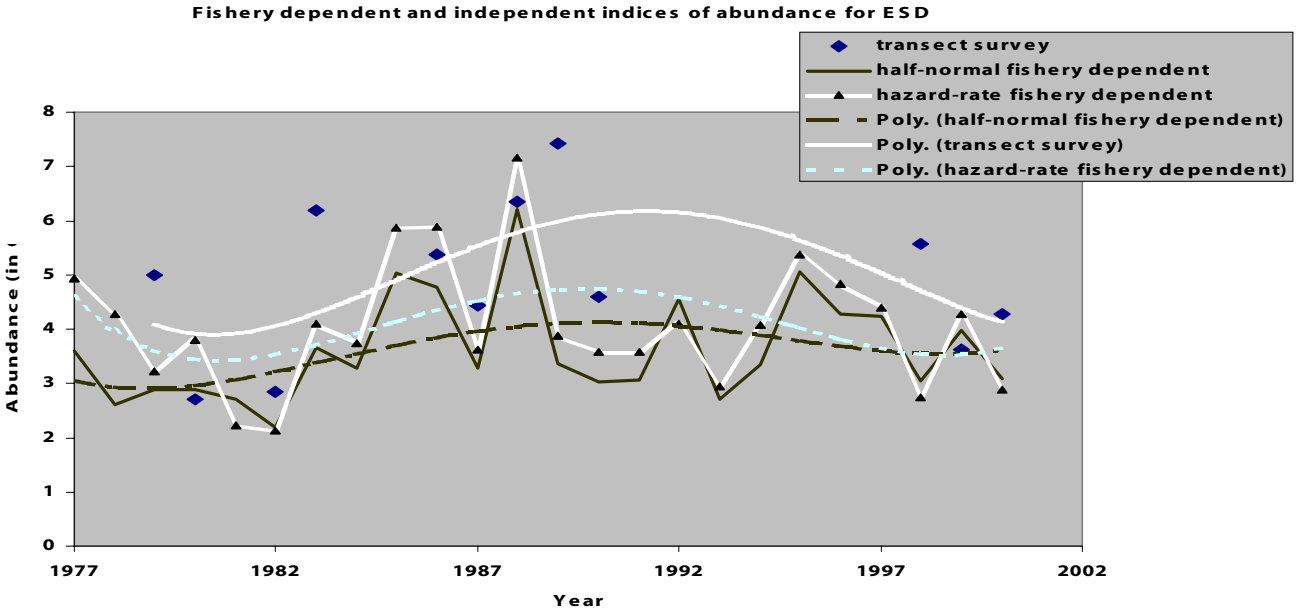
Rather than throwing out the fishery dependent data entirely, it appears reasonable to go one or two steps further to evaluate the relative merits of these data when compared against the fishery independent indices of abundance from the line transect surveys. To facilitate this evaluation, the temporal patterns in the fishery dependent time series were compared with those in the line transect series of abundance for both the NEOSD and the ESD populations. Fourth order polynomials were fitted to each time series and the patterns in the polynomials compared with each other. The fishery dependent time series for the NEOSD and ESD populations (Table 3 in Lennert-Cody et al. 2001) and the fitted fourth order polynomials were plotted (Fig. 4).

Figure 4. Plots of fishery dependent and fishery independent abundance indices and polynomial fits to the series for a. NEOSD and b. ESD.

a.



b.



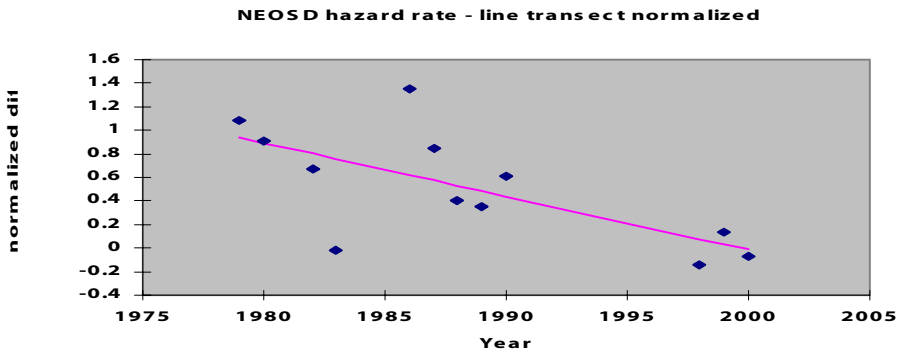
The patterns in the different series of indices are quite strong and surprisingly consistent. For data series starting in the late 1970's, the fishery dependent and fishery independent indices followed the same general patterns of a decrease from 1977, an increase from the early 1980s through to the late 1980s and a decline in the 1990s. These are the basic patterns in all of the data series and quite remarkably for both populations. There are however, some interesting differences among the indices in their time series patterns for each population. For both the ESD and NEOSD, the hazard-rate model gave a slightly larger decline than the transect survey estimates throughout the entire time series. The half-normal model series gave a lower rate of decline than the hazard rate model and the half-normal model decline was greater

than that for the line transect data for NEOSD but not for ESD. This is mainly because the half-normal model reduced the effect of decreasing proportions of sightings on and very near to the track line on the negative trend in the abundance indices. The heightened rate of decline is more strongly pronounced in the hazard rate series for the NEOSD than for the EDS.

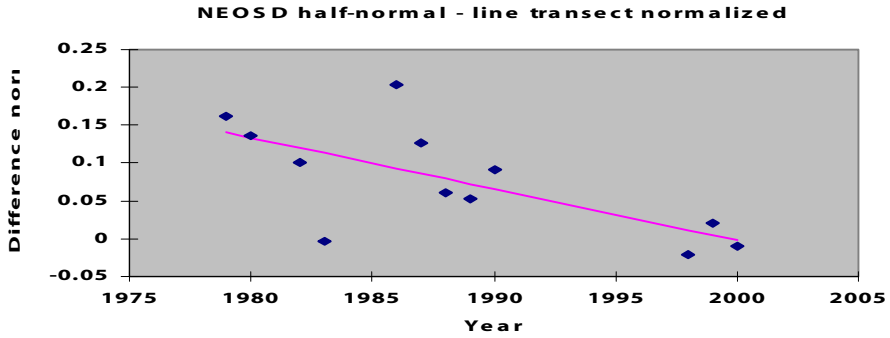
To further evaluate the differences in trends between the fishery dependent and fishery independent time series, the differences in the observations were computed for each year for the various fishery dependent and fishery independent time series. The differences were normalized by dividing the differences by the average value for the transect series. A linear regression model was fitted to each of these data sets of normalized difference versus year (Fig. 5).

Figure. 5. Plots of normalized differences between the fishery dependent and fishery independent indices. a. NEOSD hazard rate – line transect. b. NEOSD half-normal – line transect. c. ESD hazard rate – line transect. d. ESD half-normal – line transect.

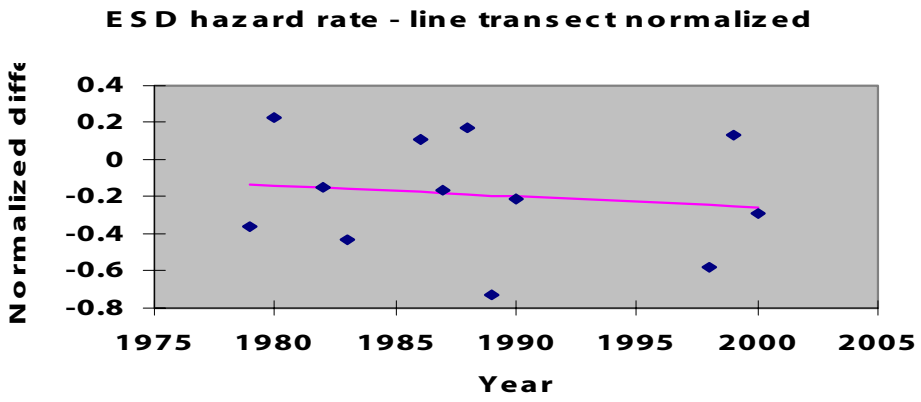
a.



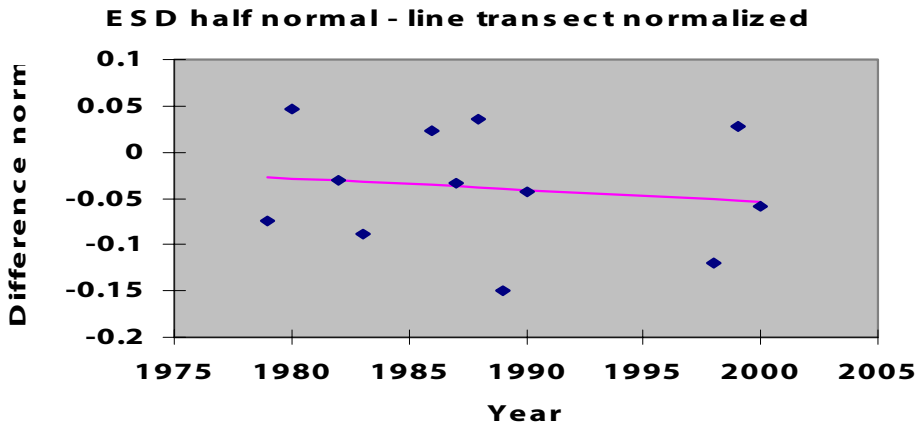
b.



c.



d.



For the hazard-rate – line transect comparison for NEOSD, a significant linear decrease was found (p-value = 0.015, n = 11). The estimate of the slope is –4.5% with a 95% confidence interval of about –1% to –8%. For the half-normal– line transect comparison for NEOSD, a significant linear decrease was also found (p-value = 0.015). The estimate of the slope is –0.7% with a 95% confidence interval of about –0.2% to –1.1%. The estimate of the slope was negative in both cases for the ESD (Fig. 5). The values were -0.6% and –0.1%, respectively, for the hazard rate – line transect and half-normal – line transect regressions. However, the slope estimates were not significant.

In summary, the regression analyses indicate that relative to the trends in the fishery independent transect estimates of abundance, the fishery dependent indices for NEOSD give negatively biased estimates of trends in abundance. The bias pattern appears to take on a linear pattern over the time period. For example the difference between the half-normal series, and line transect series changes on average by –0.7% per year when normalized by the mean line transect abundance. For ESD, this pattern is less well pronounced but still indicates a slight negative bias with the mean average normalized difference -0.1% per year for the half-normal series.

Recommendations regarding the use of fishery dependent indices of abundance

The trend-bias in fishery dependent abundance indices for ETPD can be removed using by applying expert knowledge based on Lennert-Cody et al. (2001) and empirically-based statistical methods and these datasets could thereby provide a valuable contribution to the ETPD stock assessment. Throwing

these data out without considering the development of a bias-correction for them is throwing out perfectly useful data. It is strongly recommended therefore that approaches to bias-corrections for these data be considered, that the reliability of the bias-correction approach and the potential benefits to the assessment of incorporating the bias-corrected data be simulation tested, and if found to be reliable that the bias-corrected data be incorporated into the stock assessment for ETPD. Appendix 3 outlines some suggested approaches for removing the bias in these indices and including them in the assessment together with the line transect abundance indices.

A constant linear correction for the fishery dependent index should be considered for the following reasons:

(1) **Conceptual support**. The thorough evaluations of the fishery dependent indices in Lennert-Cody et al (2001) indicated that the mechanisms most likely to cause temporal changes in bias in the fishery dependent (fd) indices were operating in a gradual manner over time, rather than at fixed and abrupt points in time. In other words, the causes of the biases were concluded to result from progressive changes in data quality and gradual fishery introduced-biases (please see the last two paragraphs on page 13 above for further details). All of the mechanisms identified to create trends in bias were found to create a negative bias that were expected to increase gradually in magnitude between 1979 and 2000. When considered to operate in combination, the combined effect of these various different mechanisms on the trend in bias should be that bias changes gradually not abruptly at certain points. The idea that ***step changes*** at either the imposition of regulatory measures that would affect reporting or at years of the major climatic regime shifts were ***not*** emphasized as being important in Lennert-Cody et al (2001). Moreover, if the latter were to occur, then the trend in the fishery independent indices would also be biased and need correction. Gradual changes in bias can be modeled by either a linear or exponential model.

(2) **Empirical evidence and parsimony**. A useful empirical diagnostic to evaluate whether a linear model is appropriate is the plot of the differences between the fishery independent (fi) and fd indices against time as indicated above (Fig. 5). In the ETPD assessment, the fi data are considered to be bias-free in their trends, and this notion is perfectly reasonable as it is in many other stock assessments. If this notion is not contested, then the above empirical comparisons between fi and fd indices for ETPD for years 1979-2000 (Fig. 5) strongly suggest that if there are temporal changes in bias in the fishery dependent (fd) indices, then these can be described adequately by a linear model over this period. In other words, on average, the absolute deviation between the true abundance and the fd increases in a linear fashion (Fig. 5) and the deviation has a negative trend. An exponential decline model could also be fitted to the data and a comparison with a linear model would require Bayes' factor for a probabilistic test of goodness of fit. However, the pattern in the differences (Fig. 5) appears to be linear and not exponential.

(3) **Feasibility of the estimation of the trend in bias**. Using the assumption of a linear model for changes in bias in the fd indices, which is supported empirically, (or an exponential decline) it would be feasible to estimate the trend in bias directly by modeling the differences between points in the fi and fd series. This would require the estimation of a single new model parameter for the slope of the trend in bias. In contrast, if a step change model was to be applied instead, and there were a number of points in the time series where changes in bias were identified, then this would require the estimation of more than one parameter (one for each step change), the degrees of freedom in the estimation of changes in bias would be much lower and the estimation of the changes in biases much less precise.

(4) The adequacy of the statistical methods, particularly the Bayesian methods, used to estimate growth rates.

Discussion of importance sampling

The sampling importance resampling (SIR) algorithm was applied to estimate posterior probability distributions for model quantities (Rubin 1988, Gelfand and Smith 1991). This algorithm is a relatively simple and reliable method to integrate joint posterior probability distributions for model parameters. If the priors are used as the importance function, it is relatively simple to check whether the method will provide an efficient approach to numerical integration. The approach used in this case was to check the maximum number of non-unique draws were obtained in the resample. A threshold value for this was set and it was found that fairly large numbers of draws from the importance function were required, yet the computing time was still very reasonable. This diagnostic appears to be quite satisfactory to check that the results obtained could reasonably be considered to be draws from the joint posterior distribution. However, if other types of importance functions are used (e.g., a multivariate log –t distribution with the mean based on the posterior mode and the covariance based on the negative inverse of the Hessian matrix (McAllister and Ianelli 1997)), other types of diagnostics should be applied. These would include testing whether the same marginal posterior distributions could be obtained when the importance function was changed slightly or moderately (e.g. if it were to be made flatter or the modes of it were shifted by some small amount e.g., 5%). Additionally, MCMC could also be applied to check to see that the same estimates were obtained.

(5) The adequacy of interpretations of the stock assessment results and protocols used for model selection when alternative models with competing explanations for historic population dynamics are fitted to the same data.

Estimation of growth rates and abundance

Abundance and growth rate estimates were similar among the alternative population dynamics models applied for each species. The estimates of r_{max} based on P. Wade's Bayesian estimation using the non-time-break models are centered over values lower than expected for ETPD (Reilly and Barlow 1986). The values for NEOSD were centered around 2% and those for ESD were around 1%. This indicates that both populations are not increasing as much as would be expected from the reduction in purse-seine kills. The estimated depletion levels are also very low for both populations, about 20% for NEOSD and 25% for ESD. The probabilistic estimation of these quantities and has been done with adequate statistical rigour. The main problem as mentioned above is the large gap in the line transect data from 1991-1997.

Criterion for model selection

In addition to fitting conventional exponential, generalized logistic and age structured models to the line transect abundance estimates, some models with breaks in the time series in key parameters were also fitted to the data. These included a 2-slope exponential model with an additional parameter that defines the year that the break occurs. Also included was a generalized logistic model in which two additional parameters were modeled, a change in carrying capacity and the breakpoint year. A second generalized logistic model was also evaluated that estimated two additional parameters, a change in Rmax and the breakpoint year.

These models were fitted to the data and the alternatives were cross-compared using an AIC criterion for model selection. The best models were identified based on the model having the highest AIC value. For NEOSD the one-slope exponential model was found to be better than the two-slope model. Also the simple generalized logistic model was found to be better than the breakpoint K and breakpoint Rmax

versions. For ESD, the two-slope exponential was selected over the one-slope exponential with a breakpoint occurring between 1990 and 1996. The Rmax and K break point models were found to be equally good and were selected over the simple generalized logistic model. The breakpoint was estimated to lie between 1989 and 1997.

The AIC is a commonly applied criterion for model selection. It is flexible and adaptable to a wide variety of estimation situations. AIC is relatively easy to compute and apply: the equation for it is quite simple and the quantities required for the equation (the maximum likelihood and the number of parameters estimated) are easy to obtain; the largest value for AIC indicates the best model when the fit of the model to the data and the number of parameters estimated are taken into account. There are however a few problems with this criterion for model selection. Firstly, by itself, a value for AIC is non-probabilistic. It is simply a number. The larger the difference between the AIC for one model and another, the better overall, the selected model. The question of how much better it is difficult to interpret.

It is conventional in statistics to make probabilistic statements based on the fits of models to data. Judgments about hypotheses and model selection are typically made based on these probability calculations. This in contrast provides a more rational basis for evaluation of the credibility of alternative hypotheses or models. When nested models are evaluated F-statistics or likelihood ratios can be calculated for comparisons between less and more complex models. This provides a probabilistic basis for model selection. However, in conventional statistics the statistics used typically rely on asymptotic assumptions and fairly large sample sizes and are based on best fitting sets of values for parameters for each alternative. This can overstate the relative difference in credibility between alternative models.

A rigorous approach to model comparison in Bayesian statistics uses Bayes' factor (Kass and Raftery 1995), the ratio of the likelihood functions of the data, integrated over all possible combinations of values for model parameters. If a prior probability can be assigned to each alternative model of interest (often set to be equal across models), then a marginal Bayesian posterior probability can be computed for each alternative model and there is no limit to the number of different models that can be compared. They must, however, all be fitted to the same dataset. Bayes' posterior for a given model gives us the probability that a model is true, relative to the others also evaluated, given the available data. If two alternative models have similar probabilities e.g., within a factor of 20 of each other, then we can say that they have similar credibility given the data. A high probability for one model, e.g., > 99% indicates that it has the highest credibility given the data. In the present case, the fits of all forms of the logistic, exponential, and age-structured models could be compared at once using Bayes' posteriors.

Bayes' factors and posteriors can be computed for alternative models using importance sampling (Kass and Raftery 1995; McAllister and Kirchner 2002). The calculation is more cumbersome than AIC and diagnostics are required to ensure that the numerical algorithm applied is converging appropriately. Wade (in press) has already applied Bayes' factors using importance sampling in previous works. Therefore, it is recommended that Bayes' posterior probabilities be applied instead to evaluate the credibility of alternative functional forms for the population dynamics models considered, particularly the breakpoint versus non-breakpoint models.

Due to the absence of data between 1991 and 1997, the statistical power of the evaluation for a breakpoint is expected to be relatively low. This is perhaps why the non-breakpoint model was found to be the best model for NEOSD. In contrast, the fishery dependent time series for each species has relatively high precision and estimates are available for all years from 1979 until 2000. It is recommended that the bias-correction procedure and the fishery dependent indices also be incorporated into the stock assessment to improve the statistical power of the tests for breakpoints.

Other alternative models that should also be considered for evaluation.

While the reported mortality of dolphins in tuna purse-seines reached over 300,000 in the 1960s, it dropped to a few thousand in 1983, increased to over 50,000 in 1986 and has dropped to about 300 animals per year in the late 1980s. Although the reported bycatch has drastically decreased, the populations are not recovering as expected (SWFSC. 1999; Wade 2002). It is therefore important that the stock assessment modeling methods applied to assess growth rates of ETPD populations include model alternatives that try to account for the lack of recovery.

"*mu-model*"

Recently, a model (called the "mu-model") for additional unaccounted mortality was evaluated in the ETPD assessment (SWFSC 1999). The model assumed that in 1992 and in later years some additional unaccounted for source of mortality began to occur. This model appeared to fit the data reasonably well compared to the other models but was dropped from the evaluations because of apparent confusions in having it understood when presented to non-scientists. This alternative model appears to be a perfectly sensible and credible model to evaluate against the data. At the meeting at the beginning of April in la Jolla, some comments were made that gave some reason to believe that in the 1990s, changes in the tuna fishery management structures, data collection, and observer programs could potentially have resulted in a decrease in reported dolphin kills in the tuna fishery. There appears to be a good case for continuing to include this model in the stock assessment as one of the alternative explanatory and predictive models for potential recent changes in population dynamics. It is recommended also that a more transparent and easy to understand name and description of this model be developed so that the idea behind it can be easily understood by non-scientists.

Due to the large gap in the data between 1991 and 1997, the power of the test for an increase in the unaccounted for mortality is low for each species. This could be remedied by increasing the amount of data relative to the number of parameters estimated. One possible implementation would be to model both the spotted and spinner dolphin stocks simultaneously as separate stocks in the same population dynamics model. Some parameters could be considered to be similar or the same between the two populations and combined into a single common parameter. An example could be the parameter expressing a mean rate of unreported mortality in the mu-model. This same approach of modeling both species in the same estimation model could be applied in other evaluations. Other parameters that could be considered to be similar or the same could include Rmax, the breakpoint year in the breakpoint models, and the relative deviation in Rmax, and K after the breakpoint, in the breakpoint models. As mentioned above, Bayes' factors and marginal posteriors could be computed to compare the relative credibility of the alternative models considered, even when populations are modeled separately or within the same model. For two separate estimations, Bayes' factor would be the product of the integrated likelihood functions for the two separate species evaluations. This would be directly comparable with the integral of the product of the likelihood functions when the two species were included in the same assessment model. It should be noted though that the estimates of NEOSD and ESD are not statistically independent because they were obtained each year from the same line transect surveys. This would need to be taken into account in a combined analysis.

Modeling calf mortality as a function of the annual per capita number of dolphin sets

The number of dolphin sets per year is important to consider because recent studies (Archer et al. 2001) have found that the chance of unaccounted calf mortality, especially, for the youngest calves may increase as a result of separation from their mothers during the vessel chase and purse-seine setting. They indicate that lactating females and calves are likely to be more vulnerable to the chase and capture process due to the energetic drain from lactation and the weaker swimming ability of calves compared to mature, non-

lactating animals. They also indicate that the mortality of calves from separation with their mothers is not counted by fisheries observers because it does not occur in the purse-seine. To evaluate the implications of this unobserved source of mortality, it is useful to evaluate the overall average number of purse-seine sets experienced by each member of a dolphin population in a year. This can be obtained by multiplying the number of sets in a year by the average number of dolphins of a given species that is set upon or chased and dividing this product by the estimated abundance of the species:

$$E_y = \frac{s_y c_y}{N_y}$$

where E_y is the per capita rate of exposure to dolphin sets in a year, s_y is the number of dolphin purse-seine sets in the year, c_y is the average number of dolphins set upon per dolphin set in a given year and N_y is the abundance of dolphins of the species of interest. c_y could also be taken to be the estimated number of dolphins chased for each set made. This calculation was first performed using the data provided on total dolphin purse-seine sets (T. Gerrodette) (Fig. 6), long-term average number of NEOSD in a set (560 from Archer et al. 2001) and the posterior median estimates of NEOSD abundance provided by a fit of the age-structured model to the line transect estimates of abundance (source: P. Wade). Provisional data were more recently provided for NEOSD on the number of NEOSD sets (Fig. 6), total NEOSD set upon and NEOSD chases. Similar calculations were also performed using these figures.

If the number for E_y is low (e.g., less than a few per year), the effect on calf mortality could be expected to be relatively low. However, if it E_y is high in at least some years and variable over time, E_y, should be a concern and considered in the population modeling. It was previously approximated that the average historic values for E_y were about 8 sets per capita. Given that the percentage of lactating females caught in sets is about 35% (Archer et al. 2001), this appears sufficient to increase the chance of calf mortality in the ETPD populations. The interannual variability in E_y, however, was not known.

At the la Jolla meeting, the reviewers requested further information about temporal patterns in fishing effort in the SE Pacific tuna purse-seine fishery. A graph of the number of pure-seine sets by type (dolphin sets, log sets, and school sets) for the years 1961-1999 was provided (Fig. 6). After this a file containing the total number of NEOSD sets for years 1977-2000 was also provided (Fig. 6). The total number of sets rose to a peak in 1979 dropped by 40% in the mid-1980s and then rose in the late 1980s and 1990s to reach an all time high in 1998. The key type of set is obviously the dolphin set. The total dolphin sets and those on NEOSD showed the same pattern. The number of the total dolphin sets increased in the 1960s and was relatively steady from 5000-8000 in the 1970s and dropped gradually to a temporary low of 3,500 in 1983. Most dramatically, the number of dolphin sets shot up to an all time high of 13,300 in 1987 and remained high for a number of years before fluctuating down to 7000 in 1993. The number back up to 10,600 in 1998.

When the per capita average annual number of sets per year was computed a similar pattern was observed (Fig. 6). The pattern computed from the total dolphin set data and 560 average number of NEOSD per set (Archer et al. 2001), was similar to those computed from the data on NEOSD sets and NEOSD chased and NEOSD set upon (Fig. 7). Using the longer total dolphin set series, this quantity averaged about 8 in the 1970s, declined to about 4 in 1983, and then shot up to about 15 sets per animal in 1987 and remained over 11-15 until 1992. This reduced to 8-9 sets per animal 1993-1996 and then climbed to 10-12 sets per animal 1997-1999. The most striking pattern is the increase from a relative low value for E_y in the mid-80s to values about four or more times higher in the early 1990s (depending on which series you look at). If unaccounted for calf mortality is a function of the annual exposure to dolphin sets, then the unaccounted for calf mortality should have increased considerably in the late 1980s and 1990s.

Because the per capita number of sets (E_y) has been large enough to conceivably increase rates of calf mortality and interannual variability in E_y is high and has increased markedly since the mid-1980s, it is recommended that the age-structured model be extended to model annual calf mortality rates as a function of the annual per capita exposure to dolphin sets. Possible model structures for calf mortality include the following.

$$M_y^1 = M_0^1 \exp\left(hE_y\right)$$

where M_y^1 is the instantaneous annual rate of total unreported mortality (natural and other) of one-year-old calves in year y, M_0^1 is the rate of total unreported mortality of calves in the absence of dolphin purse-seine sets, *h* is an estimated parameter that scales the per capita number of dolphin sets to the total unreported rate of calf mortality and E_y is the annual per capita number of dolphin purse-seine sets (defined above). The prior density for parameter *h* should be constrained to cover only non-negative values to avoid anomalous model behaviour. An alternative model form that could be considered is:

$$M_y^1 = M_0^1 + mE_y$$

where *m* is a scale parameter analogous to *h*, which should also be constrained to be non-negative. The second term on the right-hand side of the equation has the convenient interpretation of being an additional unreported rate of mortality that increases linearly with E_y. Bayes' factors and posterior could also be applied to evaluate the plausibility of these two alternative model forms. Of course, only one should be chosen for the final analysis and comparison with the simpler age structured model that models the rate of calf mortality as constant over time.

Figure 6. Estimates of the total number of tuna purse-seine sets, dolphin sets, and sets on NEOSD dolphins.

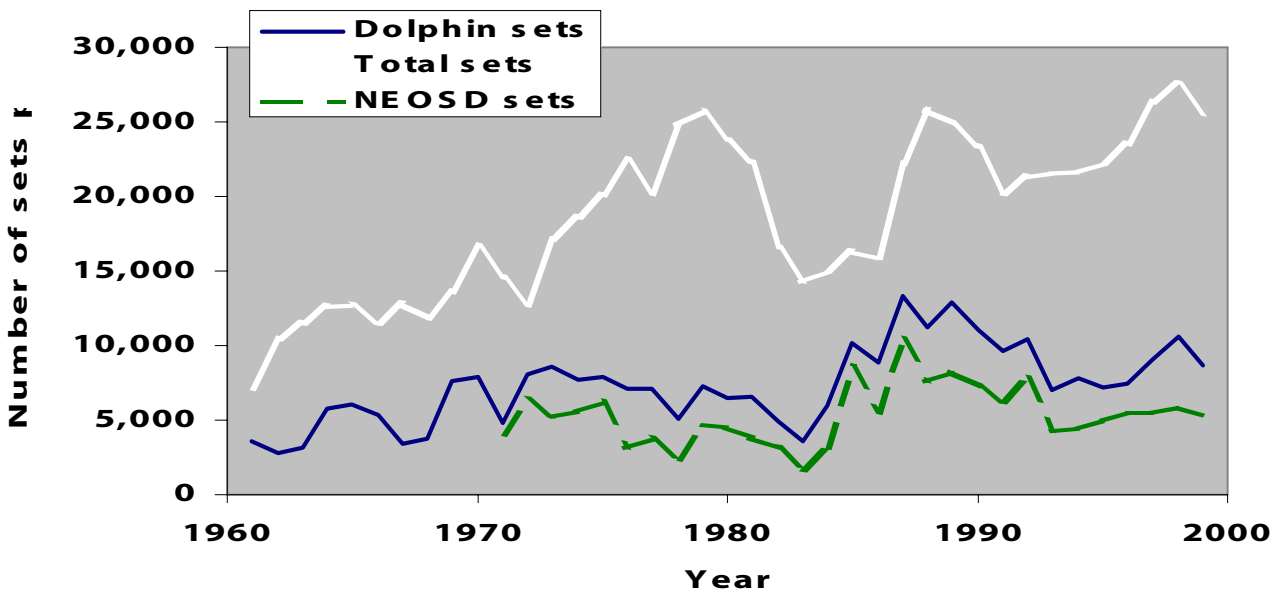
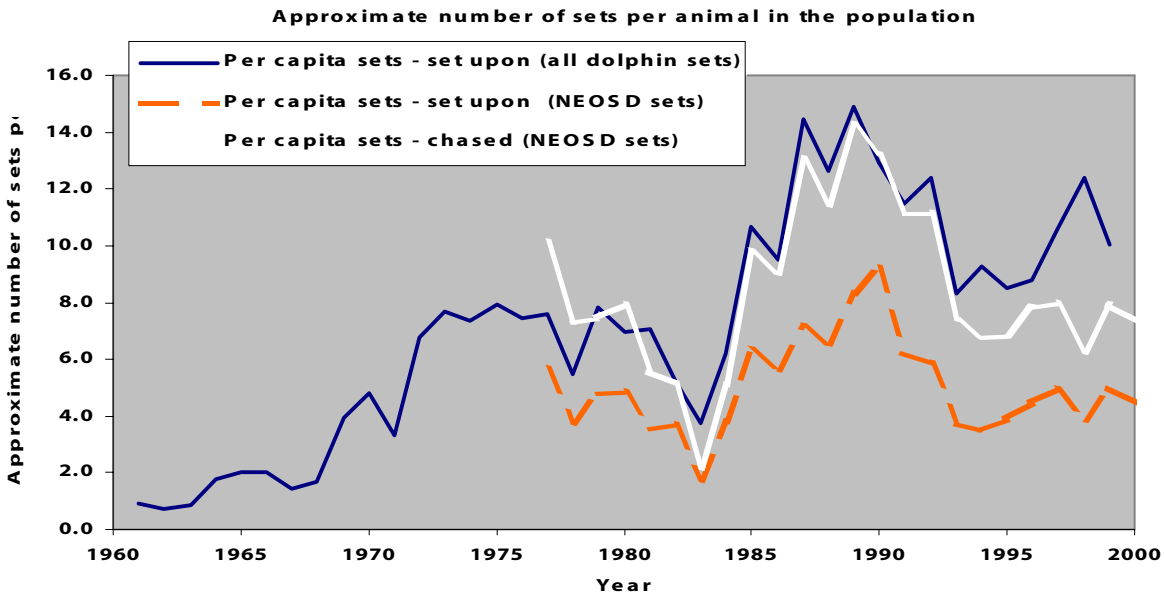


Fig. 7. Estimates of the per capita number of sets based on set upon animals using a. the total dolphin set data and the average of 560 NEOSD per set from Archer et al. (2001), b. annual estimates of NEOSD sets and estimates of the total number of NEOSD set upon, c. annual estimates of NEOSD sets and the total number of NEOSD chased. The posterior median abundance estimates for NEOSD were obtained from P. Wade.



(6) The adequacy of procedures and diagnostics used to evaluate the estimation performance of the estimators applied, the goodness of fit of models to the data and the computational performance of algorithms to compute Bayesian posterior distributions.

To evaluate the goodness of fit of models to the data, 95% probability intervals for the data were computed based on the median estimate of population abundance. This used the estimated annual sampling CV for the line transect abundance estimates. It was then evaluated how many if any of the observations fell outside of these calculated probability intervals. If only one or two observations fell outside, the model was judged to fit the data reasonably well. This approach to evaluating goodness of fit, although somewhat ad hoc, makes good sense, and should continue to be applied. However, other more formal diagnostics could and should be applied.

Gelman et al. (1995) suggest other diagnostics for model checking. These include calculating model deviance, based on a Chi-square statistic that is computed from the deviations between observations and model predictions and a measure of the variance in the observations:

$$\chi^2_{n-p} = \sum_{i=1}^{n} \frac{\left(y_i^{obs} - y_i^{pred}\right)^2}{\mathrm{var}\left(y^{obs} - y^{pred}\right)}$$

where $\mathrm{var}\left(y^{obs} - y^{pred}\right)$ is the model estimated value (of fixed value applied in the estimation) for the variance in the deviations between the observed values and model-predicted values for the observations, *n* is the number of observations, and p is the number of estimated parameters. The p-value for the Chi-square statistic is found. If p-value is less than 0.01 this implies that the chi-square value is too large and

that the assumed or estimated model variance for the time series is too small relative to the magnitude of the deviates. If the p-value is greater than 0.99, then the model variance is too large, relative to the size of the deviates.

Gelman et al. (1995) also suggest the use of the p-value of each observation as a diagnostic. This requires simulating the distribution of model-predicted observations for all observations in the dataset. The p-value or probability of obtaining a value as extreme or more extreme than the one obtained, given the model-predicted distribution of the observation should be calculated for each observation. If fraction of the p-values are very small, the model should be rejected and restructured to better predict the data. It is recommended that at least the chi-square model deviance statistic be computed as a diagnostic for goodness of fit of the model to the data.

Another approach to evaluating the estimation performance of the stock assessment methods is to randomly simulate data using a given population dynamics model, model for the data, and known values for parameters of this model. Many such datasets should be generated and then fed into the stock assessment methodology of interest. If the estimator works well the probability intervals for the parameter estimates obtained from it should overlap the true underlying values in most simulations. This approach should be applied to evaluate the estimation performance of all of the estimators already applied and the ones proposed in this review. The model used to simulate the data should be age-structured to more accurately model the data as they might actually be generated from the actual dolphin populations. Alternative underlying population dynamics models should be assumed when simulating the data to test the robustness of the estimators to variations in actual population dynamics. It would be particularly important to apply this simulation testing procedure to the proposed estimation procedure that uses both the bias-corrected fishery dependent and fishery independent estimation procedures. This would be important to evaluate whether the proposed bias-correction methodology for the fishery dependent indices works reliably in removing trend-bias from the relative abundance indices and helps to improve precision in the estimates of abundance and population parameters.

Conclusions/Recommendations

•Is the modeling approach appropriate?

Yes. The modeling approach is appropriate. The Bayesian estimation approach is arguably the most appropriate statistical approach. It is more appropriate than maximum likelihood and bootstrapping approaches that could also be applied. The use of alternative population dynamics models including exponential, generalized logistic and Leslie Matrix – age-structured models provides a robust approach to evaluating whether the population is recovering, and the potential rate of increase in abundance. The calculation of probability distributions for abundance estimates, growth rates, and the extent of depletion helps to quantify uncertainty in a transparent and easy to interpret manner. The main weakness of the approach is the lack of line transect abundance data between 1991 and 1997. Without using any other data series that would fill this gap, the ability to make sound statistical inferences about population patterns over the period of the 1990s will be limited. The inferences about these years will be based entirely on interpolation over a wide span of years. More sound statistical inference could be obtained if the fishery dependent time series, which are currently excluded, could be bias-corrected for the potential biases in the trend information contained in them and included along with the line transect abundance estimates.

•What more should be done?

(1) As pointed out in this review, statistically rigorous bias-correction methods exist to remove trend-biases in fishery dependent abundance indices and should be applied to the fishery dependent indices; the stock assessment models should thereby incorporate this bias-correction methodology and be fitted to both the bias-corrected fishery dependent data and fishery independent data at the same time.

(3) AIC should be replaced with marginal Bayes' posterior probabilities for each alternative population dynamics model that is compared with others and these probabilities used to compare the relative plausibility of alternative explanatory models for the recent observed patterns in abundance.

(3) The potential for unreported calf mortality rate to be a function of the total annual per capita exposure to tuna purse-seine chases and sets should be taken into account in the age-structured population dynamics model. The plausibility of these additional calf mortality rate models should be compared with models that do not incorporate this additional source of calf mortality. Marginal Bayes' posterior probabilities should be computed for each model alternative and used to rate the plausibility of the alternative models.

(4) The "mu-model" should be reapplied, its interpretation made more user-friendly and its plausibility evaluated against other explanatory models of recent patterns in abundance.

(5) A parametric model form for vulnerability at-age should be considered instead of the non-parametric specification for vulnerability at age.

(6) Monte Carlo simulation of datasets generated from an age-structured model with known parameter inputs should be applied to evaluate the estimation performance of the Bayesian estimation methods applied.

(7) Some additional model checking methods should be applied such as the calculation of model deviance to evaluate the goodness of fit of each model to the data.

(8) The absolute indices of abundance should be treated instead as relative abundance indices and an informative prior probability distribution should be constructed for the constant of proportionality that scales the actual abundance to the value of the relative abundance index.

•Have appropriate data been selected as input?

The basic data inputs to the population dynamics models are appropriate. However, the abundance data are sparse because of large gaps in the data, e.g., between 1991 and 1997. As mentioned above, methods to correct for trend-bias in fishery dependent abundance indices such as these exist and should be investigated and applied. To verify that the methodology for trend-bias removal is reliable, the Monte Carlo simulation approach suggested above could be applied to evaluate the potential improvements in estimation performance that could be obtained by incorporating in the stock assessment bias-corrected fishery dependent indices of abundance.

As mentioned above, indices of per capita exposure of the NEOSD and ESD populations to tuna purse-seine chases and sets should be computed and incorporated in the population modeling to account for potential increased unreported incidental mortality of calves due to recent increases in purse-seine sets on dolphins.

•What should be considered when interpreting the results?

If only the line transect time series are used in the statistical estimation, the statistical basis for inference about whether the population has been increasing is based on only three data points after a period of seven years from 1991-1997 with no abundance data. The intensive evaluation of a wide variety of hypotheses about potential mechanisms to try to account for the apparent patterns in these few data should be viewed in this particular light. A comparison of a variety of population model alternatives with so few data would be a serious overuse of the data and it could not be said that the results derived had a strong empirical basis. To improve upon the empirical basis for the analysis, the bias-corrected fishery dependent indices of abundance should be included together with the fishery independent indices and this improvement should be considered sooner rather than later because of its promise to improve the empirical basis for the assessment.

Keeping in mind the limitations of having only three data points for the most crucial part of the time series, evaluation of alternative population dynamics models should only be done after the bias-corrected fishery dependent data have been added to the analysis.

A decision analysis approach should be taken to interpret the results of stock assessments that incorporate both the fishery dependent and fishery independent data. This should involve the following steps:

(1) Formally identify of plausible alternative hypotheses or scenarios for mechanisms controlling population abundance.

(2) Construct population dynamics model components that are consistent with each alternative hypothesis specified in step 1.

(3) Fit each alternative population dynamics model formulated in step 2 to the data and computing a marginal posterior probability for each model.

(4) Use the marginal posterior probability computed step 3 as an indication of the relative credibility of results from each alternative model. The interpretation of current status of the dolphin populations and role of the tuna purse-seine fishery will depend strongly on the model fitted to the data. The marginal posterior probability for each model can be used to weight the results obtained from it when it comes to providing fisheries policy advice. Models should be considered to be similarly credible for marginal posterior probabilities within about a 20-fold difference of each other.

In summary, the main recommended improvements, including the bias-corrected fishery dependent abundance indices together with the fishery independent indices in the stock assessment, and evaluating the plausibility of purse-seine set and chase-induced calf mortality model, relative to other models, go hand-in-hand. The second requires the stronger empirical basis that the first will provide.

References

Archer, F., Gerrodette, T., Dizon, A., Abella, K., and Southern, S. 2001. Unobserved kill of nursing dolphin calves in a tuna purse-seine fishery. Mar. Mam. Sci. 17(3): 540-554.

Boyer, D. and Hampton, I. 2001. Development of acoustic techniques for assessment of orange roughy *Hoplostehus atlanticis* biomass off of Namibia, and of methods for correcting for bias. (Proc. Symposium 2000: A decade of Namibian Fisheries Science published in the South African Journal of Marine Science) 23: 223-240.

Branch, T. A. 2001. A review of orange roughy *Hoplostethus atlanticus* fisheries, estimation methods, biology and stock structure. (Proc. Symposium 2000: A decade of Namibian Fisheries Science published in the South African Journal of Marine Science) 23: 181-204.

Buckland, S.T., Anderson, D.R., Burnham, K.P. and Laake, J.L. 1993. Distance sampling. Chapman and Hall

Gilbert, D.J., 1992. A stock production modeling technique for fitting catch histories to stock index data. New Zealand Fisheries Assessment Research Document 92/15: 24p.

Geromont, H.F. and Butterworth, D.S. 2000. Possible extensions to the ADAPT VPA model applied to western north Atlantic bluefin tuna, addressing in particular the need to account for "additional variance". ICCAT Doc. SRCS/00/33.

Kass, R. E. and A. E. Raftery. 1995. Bayes factors. J. Am. Stat. Assoc. 90: 773–795.

Lennert-Cody, C.E., Buckland, S.T. and Marques, F.F.C. 2001. Trends in dolphin abundance estimated from fisheries data: a cautionary note. J. Cetacean Res. Manage. 3(3):305-319.

Lo, N. C., L. D. Jacobson and J. L. Squire. 1992. Indices of relative abundance from fish spotter data based on delta lognormal models. *Canadian Journal of Fisheries and Aquatic Sciences* 49:2515-2526.

McAllister, M.K., Babcock, E.A., and Pikitch, E.K. 2000. "Evaluating the Relative merits of alternative methods to weight different time series of abundance indices in stock assessment" International Commission for the Conservation of Atlantic Tunas, Madrid. Document SCRS/00/104.

McAllister, M. K. and J. N. Ianelli. 1997. Bayesian stock assessment using catch-age data and the Sampling/Importance Resampling Algorithm. Can. J. Fish. Aquat. Sci. 54: 284–300.

McAllister, M.K. and Kirchner, C.H. (2001) Development of Bayesian stock assessment methods for Namibian orange roughy, *Hoplostethus atlanticus*. (Proc. Symposium 2000: A decade of Namibian Fisheries Science published in the South African Journal of Marine Science) 23: 241-264.

McAllister, M.K. and G. P. Kirkwood. 1998. Bayesian stock assessment and policy evaluation: a review and example application using the logistic model. ICES J. Mar. Sci. **55**: 1031–1060.

McAllister, M.K., E. K. Pikitch, A. E. Punt and R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 51: 2673–2687.

McCullagh, P. and J. A. Nelder. *Generalized Linear Models*, 2nd edn. 1989. Chapman and Hall. New York.

Ortiz, M. and J. Cramer. 2000. Updated standardized catch rates for swordfish ages 3-10+ (*Xiphias gladius*) from the U.S. pelagic longline fleet 1982-1999. SCRS/00/144.

Punt, A. and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. Fish. 7: 35–63.

Quinn, T.J. II, and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press. New York.

Reilly, S.B. and Barlow, J. 1986. Rates of increase in dolphin population size. Fish. Bull. 84: 527-533.

Rubin, D. B. 1987. Comment on "The calculation of posterior distributions by data augmentation." J. Am. Stat. Assoc. 82: 543–546.

Rubin, D.B. 1988. Using the SIR algorithm to simulate posterior distributions. Pages 385–402 in J. M. Bernardo, M. H. Degroot, D. V. Lindley and A. M. Smith, eds. Bayesian statistics 3: Proceedings of the Third Valencia International Meeting, June 1–5, 1987. Clarendon Press, Oxford. 805 p.

SWFSC. 1999. Report to Congress on the status of eastern tropical Pacific dolphins. Available from SWFSC, P.O. Box 271, La Jolla CA 92038.

Taylor, B.L., and DeMaster, D.P. 1993. Implications of non-linear density dependence. Marine Mammal Science 9(4): 360-371.

Wade, P. 1995. Revised estimates of incidental kill of dolphins (Delphinidae) by the purse-seine tuna fishery in the eastern tropical Pacific, 1959-1972. Fish. Bull. 93: 345-354.Wade, P. 1999. A comparison of statistical methods for fitting population dynamics models to data. Marine Mammal Survey and Assessment Methods. GAmer et al. (eds) Balkema, Rotterdam, ISBN 90 5809 043 4.

Wade, P. 2002. Assessment of the population dynamics of the northeast offshore spotted and the eastern spinner dolphin through 2001. Doc. prepared for the Southwest Fisheries Science Center, National Marine Fisheries Service. La Jolla California.

Wade, P. In press. A Bayesian stock assessment of the eastern Pacific Gray whale using abundance and harvest data from 1967 to 1996. SC/48/AS8.

Appendix 1

Bibliography of All Material Provided For the Review

Archer, F., Gerrodette, T., Dizon, A., Abella, K., and Southern, S. 2001. Unobserved kill of nursing dolphin calves in a tuna purse-seine fishery. Mar. Mam. Sci. 17(3): 540-554.

Forcada, J., 2002. Multivariate methods for size-dependent detection in conventional line transect sampling. Report prepared for the Southwest Fisheries Science Center, National Marine Fisheries Service. La Jolla California.

Gerrodette, T., and Forcada, J., 2002. Estimates of abundance of dolphin stocks affected by the tuna purse-seine fishery in the eastern tropical Pacific Ocean. Report LJ-01- prepared for the Southwest Fisheries Science Center, National Marine Fisheries Service. La Jolla California.

Lennert-Cody, C.E., Buckland, S.T. and Marques, F.F.C. 2001. Trends in dolphin abundance estimated from fisheries data: a cautionary note. J. Cetacean Res. Manage. 3(3):305-319.

Perkins, P.C. and Edwards, E.F. 1998. Capture rate as a function of school size in pantropical spotted dolphins, *Stenella attenuata* , in the eastern tropical Pacific Ocean. Fish. Bull. 97: 542-554.

Reilly, S.B. and Barlow, J. 1986. Rates of increase in dolphin population size. Fish. Bull. 84: 527-533.

SWFSC. 1999. International Dolphin Conservation Act. Report to Congress. Available from SWFSC, P.O. Box 271, La Jolla CA 92038.

Taylor, B.L., and DeMaster, D.P. 1993. Implications of non-linear density dependence. Marrine Mammal Science 9(4): 360-371.

Wade, P. 1995. Revised estimates of incidental kill of dolphins (Delphinidae) by the purse-seine tuna fishery in the eastern tropical Pacific, 1959-1972. Fish. Bull. 93: 345-354.

Wade, P. 1999. A comparison of statistical methods for fitting population dynamics models to data. Marine Mammal Survey and Assessment Methods. GAmer et al. (eds) Balkema, Rotterdam, ISBN 90 5809 043 4.

Wade, P. 2002. Assessment of the population dynamics of the northeast offshore spotted and the eastern spinner dolphin through 2001. Doc. prepared for the Southwest Fisheries Science Center, National Marine Fisheries Service. La Jolla California.

Wade, P. In press. A Bayesian stock assessment of the eastern Pacific Gray whale using abundance and harvest data from 1967 to 1996. SC/48/AS8.

Appendix 2

STATEMENT OF WORK

Consulting Agreement Between The University of Miami and Dr. Murdoch McAllister

Background

The tuna purse-seine fishery has used the association between tuna and dolphins to fish in the eastern tropical Pacific Ocean for over five decades. Three stocks of dolphins were depleted by high historical levels of dolphin mortality in tuna purse-seine nets, with an estimated 4.9 million dolphins killed during the fourteen year period 1959-1972. After passage of the Marine Mammal Protection Act (MMPA) in 1972 and the increased use of fishing equipment and procedures designed to prevent dolphin deaths, mortality decreased during the late 1970s, 1980s and 1990s to levels that are generally considered biologically insignificant.

While changes in the fishery have greatly reduced the observed mortality of dolphins dramatically, the MMPA, as amended by the International Dolphin Conservation Program Act, requires that the National Marine Fisheries Service (NMFS) conduct research consisting of three years of population abundance surveys and stress studies to form the basis of a determination by the Secretary of Commerce regarding whether the "intentional deployment on, or encirclement of, dolphins by purse-seine nets is having a significant adverse impact on any depleted dolphin stock". Specific to this review, NMFS must essentially determine whether or not the depleted dolphin stocks are recovering, and if so, at what rate and at what level of certainty.

The topic of this review is the overall framework that will be used to estimate the growth rate of two dolphin populations of interest, the northeastern offshore spotted dolphin and the eastern spinner dolphin. The framework uses growth rates estimated by fitting a population model to the three-year survey estimates and other available estimates of abundance. Estimates from research vessel surveys using line transect methods are available for three periods: 1979-83 (four estimates), 1986-90 (five estimates), and 1998-2000 (three estimates), for a total of twelve estimates over twenty-one years. Reviewers will also be asked to evaluate the inclusion or exclusion of a set of fishery-dependent indices of abundance, resulting from data collected by observers onboard tuna vessels. Two types of population growth rate will be estimated: (1) exponential rate of change from 1979-2000 and (2) intrinsic rate of increase under the assumption of a density-dependent model where pre-exploitation population size in 1958 is considered carrying-capacity. Both an aggregated population model and an age-structured model will be used. Bayesian statistics, using a numerical integration method, will be used to estimate a probability distribution for the population growth rate.

Specific Reviewer Responsibilities

Expertise needed to review these analyses will include familiarity with population dynamics and assessment models, including estimation of trend and estimation of status (depletion level) using density-dependent models. Additionally, familiarity with Bayesian and likelihood estimation, including numerical methods, will be essential. Documents supplied to reviewers will include draft manuscripts, and a number of background papers (relevant publications and reports). The raw data and software used in the analysis will be made available to the reviewers upon request during the review.

The reviewer's duties shall not exceed a maximum total of two weeks, including several days to read all relevant documents, two days to attend a meeting with scientists at the NMFS La Jolla Laboratory, in San

Diego, California, and several days to produce a written report of the reviewer's comments and recommendations. It is expected that this report shall reflect the reviewer's area of expertise; therefore, no consensus opinion (or report) will be required. Specific tasks and timings are itemized below:

1. Read and become familiar with the relevant documents provided in advance;

2. Discuss relevant documents with scientists at the NMFS La Jolla Laboratory, in San Diego, CA, for 2 days, from 3-4 April 2002;

3. No later than May 3, 2002, submitting the written report[2] addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Die, via email to ddie@rsmas.miami.edu.

Signed_____ Date_____

[2] The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

ANNEX I: REPORT GENERATION AND PROCEDURAL ITEMS

1. The report should be prefaced with an executive summary of findings and/or recommendations.

2. The main body of the report should consist of a background, description of review activities, summary of findings, conclusions/recommendations, and references.

3. The report should also include as separate appendices the bibliography of all materials provided and a copy of the statement of work.

Appendix 3: Some proposals regarding the use of fishery dependent indices of abundance

In fisheries stock assessment, fisheries scientists often use relative abundance indices computed from fishery catch rate data and attempts are made to remove extraneous effects on the catch rate indices (e.g., changes in fleet composition and areas and seasons fished) that may cause them to be biased as relative indices of abundance (Lo et al. 1992; Oritz and Cramer 2000). General linear models (McCullagh and Nelder 1989) are often applied for this purpose. In other cases, independent estimates of the annual rates of change in biasing factors have been researched and are applied to remove bias in the annual trend estimates given by each relative abundance time series (e.g., Baltic herring cpue, Mika Rahikainen, pers. commn U. Helsinki).

The reasons why such efforts are made are to increase the reliability and amount of data available for fisheries stock assessment. Data on trends in relative abundance are usually very difficult and expensive to obtain. Data are typically imprecise and sparse and there are often gaps in the available time series of abundance indices. Estimates of abundance, trends in abundance and parameters such as population growth rate from sparse data can be imprecise and unreliable because no other independent sources of data are used to corroborate the trend information. The development of alternative time series of data on relative abundance can help to improve precision in estimates and can provide independent tests of assumptions on potential trends in abundance. If all available time series indicate the same trends (after bias-correction), this provides much stronger evidence for the model-estimated trend than if only one time series were to be used. If different time series show different trends and there is no method to evaluate the reliability of the different time series and estimate potential bias in the trend information given by each, then this heightens the uncertainty in the estimated trend in abundance.

Even when supposedly unbiased fishery independent indices of abundance are available, it is common to fit stock assessment models to both the fishery independent and fishery independent sources of data, as long as efforts have been taken to assure that trend-biases in relative abundance data have been removed. This is again to help the statistical power of the estimation by providing a larger amount of data for the estimation.

The sweeping argument in Lennert-Cody et al. (1995), that because the fishery independent data are trend-biased, they should not be used in stock assessment is well-intentioned and perfectly sensible. However, the advice does not take into account the variety of approaches available to correct for trend-biases in time series and the value to stock assessment of including additional time series of relative abundance indices. It is analogous to telling fisheries stock assessment scientists that because hydro-acoustic data provide biased estimates of abundance they should never be used in stock assessment. Fortunately, there is a way around this problem. By treating hydro-acoustic abundance estimates as relative indices of abundance in a stock assessment model and estimating a constant of proportionality that scales absolute abundance to the acoustic estimates, this "bias" can be estimated and the use of the "bias-corrected" data can help to improve the empirical basis for stock assessment.

By analogy, trend-bias in fishery dependent abundance indices for ETPD can also be removed using empirically-based statistical methods and these datasets could thereby provide a valuable contribution to the ETPD stock assessment. Throwing these data out without considering the development of a bias-correction for them is throwing out perfectly useful data. It is strongly recommended therefore that approaches to bias-corrections for these data be considered, that the reliability of the bias-correction approach and the potential benefits to the assessment of incorporating the bias-corrected data be simulation tested, and if found to be reliable that the bias-corrected data be incorporated into the stock assessment for ETPD.

There are a variety of approaches that could be taken to remove the bias in the fishery dependent datasets. The first is by careful inspection of the processes by which the fishery data are procured as in Lennert-Cody et al. (2001) to evaluate the potential direction and magnitude of the trend-bias. An empirical analysis of these processes can be applied to estimate the potential direction and magnitude of the trend-bias for each individual data series. For example, based on the comparison of the hazard rate and half-normal estimates they concluded that this indicated a decrease of 0.6% per year resulting from changes in data quality. (In contrast, a regression analysis similar to that above but comparing the hazard rate and half-normal indices for the time period between 1977 and 2000 gave a significant estimated annual decrease of -1.5% per year for NEOSD and –1% for ESD.) Based on the increase in the percentage of sightings that led to sets, Lennert-Cody et al. (2001) conclude that changes in selective reporting contributed a decrease of "roughly 0.3% per year (this assumes no effect on herd size or detection near the trackline)." The combined effect based on Lennert-Cody et al. (2001) is –0.9% per year if the hazard rate model were to be used. For the NEOSD, a comparison of the hazard rate model and modes of search index which they believe to be less biased, lead to an estimate of –1.5% per year in the hazard rate index. (In contrast, a regression analysis similar to that above but comparing the hazard rate and modes of search index for the time period between 1984 and 2000 gave an significant estimated annual artificial decrease for the hazard rate model of -4.2% per year for NEOSD. A regression comparing the half-normal and modes of search indices for NEOSD found no significant difference in the trends given by them).

In summary, Lennert-Cody et al. (2001) provide approximations of trend-bias mainly in the hazard rate index assuming that the half-normal and modes of search indices are less biased. The regression analyses above indicate that this is a fair assumption for ESD. However, the regression analysis suggested that when the half-normal estimates for NEOSD were compared to the line transect estimates, there was still a –0.7% bias. The authors do, however, suggest an additional source of bias of –0.3% per year due to changes in selective reporting.

These analyses together with the ones I have done in addition provide empirical estimates of the potential trend-biases in the fishery dependent indices for NEOSD and ESD. They suggest that the half-normal model indices for ESD are less biased than the hazard rate model indices and that for NEOSD, the half-normal and modes of search indices give very similar estimates of trends. Based on the comparison of the half-normal and line transect indices, for NEOSD the trend-bias in the half-normal index was about –0.7% per year (SE=0.2%). For ESD, the trend-bias was –0.1% per year (SE=0.3%).

A suggested approach for modeling the trend-bias in fishery dependent abundance indices.

If the fishery dependent indices of abundance were to be included in the stock assessments of ETPD populations, the stock assessment methodology would need to be modified to estimate the trend-bias in these time series. Within the Bayesian approach applied, the trend-bias could be modeled in a simple yet rigorous manner. The above analysis already gives clues about the structural form of the trend-bias. It appears that a linear additive model for trend-bias could describe the trend-bias adequately.

$$b_y = 1 + a \times (y - y_0)$$

where b_y is the bias factor in year y, a is the annual proportional increase in bias which is assumed to be constant over the time series, and y_0 is the first year in which both series have an observation. The predicted fishery-dependent index in a given year would then be calculated as:

$$I_y = q \times b_y \times N_y$$

where I_y is the fishery dependent index of abundance, and N_y is the population abundance in year y. The same type of likelihood function could be applied to the fishery dependent indices I_y as has been used for the fishery independent indices, a lognormal density function. A value for the lognormal SD term, σ, would also be needed.

The issue of assigning a value for σ is contentious when different time series suggest different population trends (McAllister and Babcock 2000). This is because the value assumed or estimated for σ determines the relative weightings given to the different time series and this determines the final estimate of the abundance trends. However, when a stock assessment model is fitted to different time series, the data should be modeled such that the same population trend is being tracked by each series, and all potentially biased series should be bias-corrected so that there should be no substantial differences among the population trends implied by the different series. Thus, with the adoption of an appropriate bias-correction factor for the fishery dependent indices, the issue of the choice of a value for σ becomes less of an issue. However, to ensure stable estimation behaviour, σ for the bias-corrected indices should be no smaller than for the series considered to be unbiased. This will prevent the more complete bias-corrected series from obtaining unduly small values for σ and dominating the estimation. One potential implementation could apply the same value for σ across the different series and include an additional variance term for the bias-corrected series (Geromont and Butterworth 2000):

$$\hat{\sigma}^2_{fi,year} = \hat{\sigma}^2_{common} + \sigma^2_{sampling\ \ fi,year}$$

$$\hat{\sigma}^2_{bias\ corrected\ \ fd,year} = \hat{\sigma}^2_{common} + \hat{\sigma}^2_{added} + \sigma^2_{sampling\ \ fd,year}$$

where $\hat{\sigma}^2_{fi,year}$ is the variance for the fi series that is considered to be bias-free, $\hat{\sigma}^2_{common}$ is the estimated variance over and above the sampling variance for the most reliable dataset, the fi indices, $\sigma^2_{sampling\ \ fi,year}$ and $\sigma^2_{sampling\ \ fd,year}$ are the bootstrapped sampling variance for the fishery independent and fishery dependent indices in that year, $\hat{\sigma}^2_{bias\ corrected\ \ fd,year}$ is the variance for the bias-corrected fd time series, and $\hat{\sigma}^2_{added}$ is an additional variance term that is the be estimated based on the fit of the stock assessment model to the combined data series. Priors would need to be specified for the estimated variance terms on the right hand side of the equations so that they were centered over plausible values, non-zero and non-negative. The prior for $\hat{\sigma}^2_{common}$ apart from covering non-negative and non-zero values, should be relatively non-informative. The prior for $\hat{\sigma}^2_{added}$ should be informative and centered over a positive value at least as large as the positive difference between the mean of the $\sigma^2_{sampling,year}$ for the line transect estimates minus the mean of the $\sigma^2_{sampling,year}$ for the fishery dependent indices.

If bias-corrected fishery-dependent indices are to be considered for stock assessment, prior density functions would be required for the parameters q and a. A non-informative prior could be used for q and a. For q this would be uniform over the natural logarithm of q. For a this would be uniform over a.

Alternatively an informative prior could be applied for a, if there were expert knowledge about the potential direction and magnitude of a that were independent of analysis of the time series of observations of I_y. For this reason, the above analyses and the analyses in Lennert-Cody et al. (2001) that used the time

series of I_y could not be used to develop this prior. However, at least one other finding in Lennert-Cody et al. (2001) could be used to help construct an informative prior for *a*. They concluded that changes in selective reporting contributed a decrease of "roughly 0.3% per year (this assumes no effect on herd size or detection near the trackline)." The use of half-normal indices would partially correct for the effect on detection near the trackline. The herd size effects should also act in the same direction since larger herds tended to be reported more in later years, indicating that the rate of reporting of smaller herds decreased. Due to there being no quantitative figure being placed on the latter effect, it could be provisionally assumed to be negligible and if not, the prior for *a* should be empirically updated in the stock assessment. A provisional proposed prior for *a* for each half-normal fd series could then be centered about the value of –0.003 per year.

The prior SD for *a* could be determined by expert judgment. It could be asked without comparing the fishery-dependent and fishery independent indices available how much a fishery dependent trend could deviate from the true trend due to additional unaccounted for trend-biasing processes. It would seem that the maximum conceivable annual rate of bias could be up to 2%. Doubling this would give about 4% per year. Over a thirty-year time frame this would give rise to a 120% difference, which is perhaps as large as any difference seen when comparing fishery dependent and fishery independent data series for individual stocks. A normal prior density function could be chosen since *a* can be above or below the most credible prior value and could conceivably be either negative or positive. A proposed prior for a would thus be:

$a \sim \text{Normal}(-0.003, 0.04^2)$.

Note that this is only a suggested provisional prior for *a*; the choice of this prior, should the approach be adopted, would require input from, and careful inspection by, the biologists and fisheries scientists.

The revised stock assessment procedure would then require the population dynamics models to be fitted to both the fishery independent and fishery dependent indices at the same time. In the existing stock assessment, it has been assumed that the fishery independent indices provide unbiased estimates of the population trends. This appears to be a fair assumption and there does not appear to be a reason to change it. By assuming that the fishery independent indices are unbiased, this would enable the bias term *a* for the fishery independent indices to be estimated. The data should provide precise posterior estimation of *a* providing that the SE in the empirically estimated value for *a* is not larger than the prior SD for *a*. This latter condition regarding the information requirements for posterior updating of *a* is based on the analogy of a normal conjugate prior for *a* which incorporates a normal likelihood function; here the posterior variance is the harmonic mean of the prior and empirical variance.

$$\sigma^2_{posterior} = \frac{1}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{data}}}.$$

If the prior and data variances are equal, the posterior variance is half of the prior variance. If the data variance is one quarter of the prior variance, then the posterior variance is 20% of the prior variance. The regression analyses reported above for both NEOSD and ESD confirm that the data should enable precise estimation of the bias factor *a*. The SE's in the regression estimates of *a* were 0.002, and 0.003 for NEOSD and ESD, respectively. This would give posterior variances of approximately 0.0025, and 0.0055, respectively for the parameter *a*. It is expected that there would be a negative posterior correlation between the parameter *a* and the estimated rate of increase. However, the fishery dependent datasets appear to be large and sufficiently informative to keep this correlation small and non-influential.

As noted above, when polynomial functions were fitted to the line transect and then the fishery dependent indices they appeared to be tracking the same temporal trends throughout the 1980s and 1990s (Fig. 4). Overall, the half-normal time series should be the best fd series to work with because the trend-biases are smallest for these series and they are the longest. For NEOSD there was a consistently increasing deviation downwards relative to the fishery independent series. A significant average negative annual change in the deviation was confirmed in the regression analysis. However, for ESD there was no apparent consistent change in the annual deviation of the half-normal series and the regression analysis also confirmed no significant average annual change in the deviation. My analysis thus indicates that a bias-correction factor could be precisely estimated from a joint analysis of the fishery independent and fishery dependent data within this stock assessment procedure. A precisely estimated bias-correction function for the fishery dependent data and the inclusion of these bias-corrected data together with the line transect abundance estimates in the stock assessment should then increase considerably the precision and reliability in the evaluation of the rates of population growth for both NEOSD and ESD.

Appendix 4

Information Provided by NMFS by e-mail After the Review Meeting

Date: Mon, 22 Apr 2002 15:04:24 -0700
From: "Nicole Le Boeuf" <Nicole.Leboeuf@noaa.gov>
X-Mailer: Mozilla 4.76 [en] (Win98; U)
X-Accept-Language: en
To: Murdoch McAllister <m.mcallister@ic.ac.uk>,
Malcolm.Haddon@dpiwe.tas.gov.au
CC: Manoj Shivlani <mshivlani@rsmas.miami.edu>
Subject: Re: Assessment Model Review

Dear Murdoch and Malcolm,

I'm having a hard time catching Paul Wade as he is currently on travel, so I hope that doesn't cause too much trouble. After you see what's attached, please let me know what you still lack.

It's Steve's presentation, and the attendees were (roughly as they were coming in and out):

Dr. Paul Wade - presenter of assessment model
Dr. Steve Reilly - IDCPA research coordinator
Dr. Tim Gerrodette - generated the dolphin abundance estimates
Dr. Wayne Perryman - dolphin photogrammetry specialist
Dr. Paul Fiedler - ecosystem specialist
Dr. Lisa Ballance - ecosystem specialist
Dr. Bill Perrin - ETP dolphin specialist
Dave Bratten - IATTC

Hope that helps, Nicole

Murdoch McAllister wrote:
>
> Dear Nicole,
>
> It was nice to meet you in La Jolla and I enjoyed taking part in the two
> days of meetings at the SW lab. I have got back from my trip to Riga on
> Monday and am just now catching up with my large "in tray". I would be
> very grateful if you could arrange to send by e-mail at your earliest
> convenience the following:
>
> (1) A list of all of the names and positions of the scientists who attended
> each of the two days meetings.
> (2) File copies of the overhead, MS word and power point presentations made
> during the two days meetings.
>
> Best wishes,
> Murdoch McAllister

IDCPA Apr 02 overview.ppt

Date: Tue, 23 Apr 2002 11:18:36 -0700
From: "Nicole Le Boeuf" <Nicole.Leboeuf@noaa.gov>
X-Mailer: Mozilla 4.76 [en] (Win98; U)
X-Accept-Language: en
To: m.mcallister@ic.ac.uk, Malcolm.Haddon@dpiwe.tas.gov.au
CC: mshivlani@rsmas.miami.edu
Subject: [Fwd: [Fwd: request for some more figures]]

Here's a little more of what you requested. I will do what I can to
track Paul down for the rest.

Return-Path: <Tim.Gerrodette@noaa.gov>
Received: from sunfish.noaa.gov ([199.105.13.2]) by
HQMail.nmfs.noaa.gov (Netscape Messaging Server 4.15) with ESMTP
id GUZSLB00.062 for <nicole.leboeuf@noaa.gov>; Mon, 22 Apr 2002
19:15:11 -0400
Received: from noaa.gov ([199.105.14.77]) by sunfish.noaa.gov
(Netscape Messaging Server 4.15) with ESMTP id GUZSQM00.066 for
<Nicole.Leboeuf@noaa.gov>; Mon, 22 Apr 2002 16:18:22 -0700
Message-ID: <3CC49A37.D70ABF94@noaa.gov>
Date: Mon, 22 Apr 2002 16:18:15 -0700
From: "Tim Gerrodette" <Tim.Gerrodette@noaa.gov>
X-Mailer: Mozilla 4.75 [en] (Win98; U)
X-Accept-Language: en
MIME-Version: 1.0
To: Nicole Le Boeuf <Nicole.Leboeuf@noaa.gov>
Subject: Re: [Fwd: request for some more figures]
References: <3CC48919.843088B1@noaa.gov>
Content-Type: multipart/mixed;
boundary="----------B79A8EF40B20D08674B584EC"

Nicole, excel file for request #1 is attached. Paul Wade will send you files
for requests #2 and #3. -- Tim

Nicole Le Boeuf wrote:

> Help! Do you two have this stuff?
>
> ----------
>
> Subject: request for some more figures
> Date: Mon, 22 Apr 2002 14:26:54 +0100
> From: Murdoch McAllister <m.mcallister@ic.ac.uk>
> To: "Nicole Le Boeuf" <Nicole.Leboeuf@noaa.gov>,
> Manoj Shivlani <mshivlani@rsmas.miami.edu>

> CC: Malcolm.Haddon@dpiwe.tas.gov.au
> References: <00fa01c1d687$c0d38060$4768ab81@rsmas.miami.edu>
>
> Dear Nicole,
>
> I would be grateful if you could obtain and pass on to me a few more
> figures. These include
>
> (1) An Excel file containing the Purse-Seine Sets by Type from the early
> 1960's until 2000. A figure of this was passed out at the meeting.
> (2) Files containing the input data to the age-structured population
> dynamics models for the spotted and spinner dolphins. This would include
> for example the data for the catch at age that are used to estimate the
> vulnerability at age.
> (3) The Excel spreadsheets in which the probability intervals are
> calculated for population abundance and for the survey abundance data.
>
> Thanks,
> Murdoch
>

sets.xls

Date: Thu, 25 Apr 2002 13:49:41 -0700
From: "Nicole Le Boeuf" <Nicole.Leboeuf@noaa.gov>
X-Mailer: Mozilla 4.76 [en] (Win98; U)
X-Accept-Language: en
To: m.mcallister@ic.ac.uk, Malcolm.Haddon@dpiwe.tas.gov.au
CC: mshivlani@rsmas.miami.edu
Subject: [Fwd: data]

Greetings,

I hope this satisfies your request. Please let us know if you need
anything else.

Cheers, Nicole

Return-Path: <paul.wade@noaa.gov>
Received: from mercury.akctr.noaa.gov ([161.55.120.130]) by
HQMail.nmfs.noaa.gov (Netscape Messaging Server 4.15) with ESMTP
id GV3YH000.KTU for <nicole.leboeuf@noaa.gov>; Thu, 25 Apr 2002
01:12:36 -0400
Received: from noaa.gov ([161.55.176.51]) by
mercury.akctr.noaa.gov (Netscape Messaging Server 4.15 mercury
Jun 21 2001 23:53:48) with ESMTP id GV3YMC00.BT4; Wed, 24 Apr
2002 22:15:48 -0700
Message-ID: <3CC79140.5FBE7CAE@noaa.gov>

Date: Wed, 24 Apr 2002 22:16:48 -0700
From: Paul Wade <paul.wade@noaa.gov>
Organization: National Marine Mammal Laboratory
X-Mailer: Mozilla 4.7 [en] (WinNT; U)
X-Accept-Language: en
MIME-Version: 1.0
To: Tim Gerrodette <Tim.Gerrodette@noaa.gov>
CC: Nicole Le Boeuf <Nicole.Leboeuf@noaa.gov>
Subject: Re: data
References: <3CC49363.90BC3CFC@noaa.gov>
Content-Type: multipart/mixed;
 boundary="----------A00E67CBDBAAAF3C9183CA90"

Tim,
Input file of data and prior distributions in NE_Spotted_DataInput.inp

Age distribution of kill (smoothed) in ageselect.xls, in sheet
Data&Selectivity, under "Observed"

Results trajectory with sampling limits in ES2002_ADDNA_Send.xls

Remind Murdoch that I discovered the problem with the sampling interval
calculation -- I had changed a quick and dirty normal interval to the proper
log-normal interval, but neglected to carry it down to the last 3 abundance
estimates. I fixed it in this file at the meeting. They are properly
asymetric the correct direction, highlighted in yellow (with difference in
blue).

Cheers,
Paul
Tim Gerrodette wrote:

> Dear Nicole,
>
> I would be grateful if you could obtain and pass on to me a few more
> figures. These include
>
> (1) An Excel file containing the Purse-Seine Sets by Type from the early
>
> 1960's until 2000. A figure of this was passed out at the meeting.
> (2) Files containing the input data to the age-structured population
> dynamics models for the spotted and spinner dolphins. This would
> include
> for example the data for the catch at age that are used to estimate the
> vulnerability at age.
> (3) The Excel spreadsheets in which the probability intervals are
> calculated for population abundance and for the survey abundance data.
>
> Thanks,
> Murdoch

ES2002_ADDNA_Send.xls

NE_Spotted_DataInput.inp

ageselect.xls

Date: Mon, 29 Apr 2002 12:12:56 -0700
From: "Nicole Le Boeuf" <Nicole.Leboeuf@noaa.gov>
X-Mailer: Mozilla 4.76 [en] (Win98; U)
X-Accept-Language: en
To: Murdoch McAllister <m.mcallister@ic.ac.uk>
Subject: Re: Assessment Model Review

Murdoch,

I spoke with Tim Gerrodette about this request, and he indicated that he
doesn't have the data that you need right now, but is working on it and
will have it soon. I'll send it along as soon as it is in hand.

All the best, Nicole

Murdoch McAllister wrote:
>
> Dear Nicole,
>
> I have another request. I would be grateful if you could obtain and send
> by e-mail to me an Excel file containing the current best estimates of the
> average annual number of NE offshore spotted dolphins that were set upon
> per tuna purse-seine set for years from 1958 - 2000.
>
> Thanks,
> Murdoch
Date: Mon, 29 Apr 2002 16:29:34 -0700
From: "Tim Gerrodette" <Tim.Gerrodette@noaa.gov>
X-Mailer: Mozilla 4.75 [en] (Win98; U)
X-Accept-Language: en
To: m.mcallister@ic.ac.uk
CC: Nicole Le Boeuf <Nicole.Leboeuf@noaa.gov>,
"REILLY Steve" <Steve.Reilly@noaa.gov>,
"WADE Paul" <paul.wade@noaa.gov>
Subject: assessment model data request

Dear Murdoch,

Nicole forwarded your latest request to me regarding the number of NE
spotted dolphins set on each year. Nicole is in the midst of clearing
out her office in La Jolla and moving back to Washington, so I will

reply directly. I have attached a data set that will partially answer your request. Some caveats: First, we're still working on these numbers ourselves, so they are by no means final. Second, I have included 3 time series for NE spotted dolphins - for # sets, # individuals captured (set on), and # individuals chased. These are highly correlated but you might want to use any of them. Third, the time-series only goes back to 1971 at present, when the observer program began, and the # captured only begins in 1977 when number in the net was recorded separately. I hope these will suffice for your modelling. I'm sorry to send you preliminary numbers, but I thought it better to respond quickly rather than wait. We'll be happy to send revised numbers in a couple of weeks, but I expect you'll be done with the review by then. In any case the final numbers should be approximately proportional to these. We look forward to receiving your review.

Tim Gerrodette

NE spotted.xls
Date: Wed, 01 May 2002 08:40:26 -0700
From: "Tim Gerrodette" <Tim.Gerrodette@noaa.gov>
X-Mailer: Mozilla 4.75 [en] (Win98; U)
X-Accept-Language: en
To: Murdoch McAllister <m.mcallister@ic.ac.uk>
CC: mshivlani@rsmas.miami.edu,
Nicole Le Boeuf <Nicole.Leboeuf@noaa.gov>
Subject: Re: Fwd: Re: Assessment Model Review

Murdoch,

Meghan Donahue drove you to your hotel.
Your other names:
Dr. Bill Fox, chief scientist of one branch of NMFS (Washington)
Dr. Eric Archer - spoke about cow-calf separation during fishing
Mr Paul Fiddler - Dr. Paul Fiedler, oceanographer in previous list
Ms. Nicole le Boeuf, NMFS Washington
Ms. Meghan Donahue, SWFSC
Bob? Dr. Robert Brownell, head of Protected Resources Division at SWFSC might have been there (I don't recall).
Dr. Michael Tillman, Director of SWFSC, was also at the meeting

Tim

Murdoch McAllister wrote:

> Dear Tim,
>
> Thanks for sending the file with the new data on NEOSD sets. I sent the
> message below to Nicole but got no reply perhaps because she is in the

> middle of moving office. Is there anyone at the lab who could correct my
> notes on who attended the review meetings in early April?
>
> Thanks,
> Murdoch
>
> >Date: Tue, 30 Apr 2002 12:42:27 +0100
> >To: "Nicole Le Boeuf" <Nicole.Leboeuf@noaa.gov>,
> >Malcolm.Haddon@dpiwe.tas.gov.au
> >From: Murdoch McAllister <m.mcallister>
> >Subject: Re: Assessment Model Review
> >Cc: Manoj Shivlani <mshivlani@rsmas.miami.edu>
> >
> >Dear Nicole,
> >
> >In addition to these names, I caught partial and incorrect names of others
> >attending in my notes. I would be grateful if you could correct my
> >additions below and indicate the positions of each.
> >
> >You had provided:
> >
> >At 03:04 PM 4/22/02 -0700, Nicole Le Boeuf wrote:
> >>Dr. Paul Wade - presenter of assessment model
> >>Dr. Steve Reilly - IDCPA research coordinator
> >>Dr. Tim Gerrodette - generated the dolphin abundance estimates
> >>Dr. Wayne Perryman - dolphin photogrammetry specialist
> >>Dr. Paul Fiedler - ecosystem specialist
> >>Dr. Lisa Ballance - ecosystem specialist
> >>Dr. Bill Perrin - ETP dolphin specialist
> >>Dave Bratten - IATTC
> >
> >
> >In addition, my notes had:
> >
> >Dr. Bill Fox chief scientist, NMFS
> >Mr Eric Archer
> >Mr Paul Fiddler
> >Ms. Nicole le Boeuf
> >
> >Only first names in my notes (might be incorrect):
> >
> >Bob?
> >Megan?
> >
> >What was the name of the kind woman who drove Malcolm and I out from the
> >lab after the meeting on the second day?
> >
> >Thanks
> >Murdoch
> >