The National Marine Fisheries Service's Bottlenose Dolphin Science

An Independent Review carried out for the

Centre of Independent Experts

February 2002

Michael C.S. Kingsley

Nuuk, Greenland

CONTENTS

Preamble

It is my understanding that the intention of this review is to ensure that the scientific assessment work reported here will withstand careful scrutiny, and to point out places where this is in doubt. This does not give the same opportunity as in the usual scientific review process to commend originality, for example, and this review will probably appear more negative than the reviewer would prefer. Much of the work reported here is estimable and of scientific value, but it is not the understood purpose of this review to point it out.

Furthermore, it is my understanding is that the intention of this process is to seek an independent review. Allowance should therefore be made for the fact that ways of thinking that are so well accepted within the NMFS that their validity is not often questioned may be subjected to a different point of view, and different conclusions drawn about them. The reviewer may offer critical reviews of particular analytical (or other) methods or practices if he has doubts about their scientific validity, even though they may be common form within NMFS (or elsewhere).

General Assessment:

The scientific programme that has generated the assessment data could perhaps have produced more definite results with better planning. For example, a pilot study of the inshore-offshore distribution of genotypes might have allowed a more intense subsequent sampling programme to be concentrated at a distance offshore that would have more accurately defined the boundary between the stocks; earlier analysis of north-south structure within the coastal stocks might have allowed better planning of aerial survey programmes; and so forth. The existing research programme structure has imposed on the scientists concerned the need in some cases to cobble together estimates and analyses by combining results that were not originally intended to be combined, or were undertaken under mistaken impressions of stock structure and movements. Their performance in doing so has been creditable, but often limited as to the information that it has been possible to distil from the available data.

Given these limitations, the qualitative science is generally sound. The methods by which most problems are tackled are sensible and make best use of the information available. The conclusions about north-south stock structure are reasonable and convincingly presented. Given the existence of state jurisdiction in some sea fisheries, some state-based collection of fishery statistics, and the likely need to enlist the states in take reduction measures, the division into management units based on state lines seems like a practical approach, even if it does not correspond exactly to unit stocks. In any case, unit stocks may remain to be exactly defined and might in any case be labile over time. Dividing a coastal population into stocks affects calculations of PBR; I discuss this below.

I have not reviewed very thoroughly the genetic and other evidence presented for north-south divisions between stocks. These results seem to be still rather tentative. If divisions of this kind became better established, they could eventually lead to requirements for establishing allowable takes *by stock*. Taking such decisions would also require estimates of numbers for each stock. The information on stock size is still not in a state that would allow such decisions to be made with any reliability.

The quantitative outputs are marred by what I would have to describe as a systematic denial of uncertainty. The analyses presented give abundant evidence of enthusiasm, hard work, and immense amounts of calculation. Complex models are developed and analysed, and estimates are produced and presented, each accompanied by an estimated ECV. However, it appears that the ECVs are uniformly too low, sometimes by quite large factors. The overall effect is to give an impression of confidence and exact knowledge to results that are in truth somewhat less than that. There are several sources of this bias, not all applying in every case:

a) Unreliability in data is not always adequately treated. For example, in estimating by-catch, and as the ASRG points out, the sampling method (selection of trips to place observers on) is not mentioned or discussed, nor is the reliability of the landings data used as a measure of effort to gross up the by-catch rates estimated from observed occurrences to a total number.
b) Partial bootstrapping. If a bootstrap is to be used to estimate uncertainty of an estimate, all steps of the estimation process need to be bootstrapped, including model selection and parameter estimation (if applicable) as well as in the application of the model. It sometimes occurs that an investigator bootstraps only part of an estimation process, thus producing a small standard error while truthfully asserting that 'a bootstrap was used to estimate the error variance'. In some places in these documents, bootstrapping was used to estimate an ECV with little explanation of what distribution was built, how it was sampled, and which estimation steps were repeated on the bootstrapped samples.
c) Ignoring model selection error. Selecting, from among several models, that which can best be fitted to the data, and evaluating the resulting estimates only from the standard errors of the parameters of this best-fitting model, reliably underestimates uncertainty. In the form of 'posterior stratification', this error is well known and usually avoided; it is recognised as a problem—but usually ignored—in stepwise inclusion of independent variables in models. But in less familiar guise it is frequently committed, for example in line transect analysis (where the ECV of esw is often underestimated because the same data is used both to select and to fit a visibility model).
d) Ignoring other components of error. There are frequent statements within these documents that bias of one kind or another exists, but followed nonetheless by an estimate accompanied by an ECV which ignores the bias. For example, the uncertainty attached to aerial survey estimates of numbers seems to have no component of error to allow for the uncertainty of the correct boundary for obtaining an unbiased estimate of the coastal stock.

This problem of understated uncertainty, in the form of quantitative ECVs that are frequently too small, might be reduced if the investigators providing these reports were encouraged to state explicitly the assumptions made in producing both the estimates of central tendency and the estimates of uncertainty. It would also help if the sources of uncertainty associated with the estimates were listed, along with statements about which sources of uncertainty were included in the ECV, which were too small to worry about, and which were possibly large, but not included.

Some questionable statements in these documents could be avoided if the analytical efforts were backed up by better judgement and an improved knowledge of basic statistics. Three examples: a) the ECV of an estimate of total by-catch based on 11 observed occurrences is given as 16%. ECVs of such estimates can be expected to be governed at base by the rule of $1/\sqrt{n}$, in this case leading to an expected ECV of *at least* 30% (given other uncertainties, probably more, rather then less). A value of 16% is not believable, but it is presented without comment or discussion. b) in using strip surveys to extend line-transect estimates, the ECV of a ratio appears to have been *guessed* to be the ratio of the ECVs of the terms. Anyone carrying out analyses of this kind should be able to write down *from memory* simple expressions for combining variances; and should not even consider replacing knowledge with guesswork. c) discussing by-catch estimation, both Palka and Rossman's document and the ASRG (!) comments lend credence to the fallacy that low levels of observer coverage can bias estimates by being 'too low to of seen an observed take' (*sic*); 'due to the low coverage of a small fishery, it is likely the by-catch estimate is biased low'. No, no, and again no. Unbiased observation yields unbiased estimates; low levels of observation produce imprecision, not bias; and any elementary statistical textbook will say as much. This reasoning confuses a high probability of (small) negative error with negative bias in estimation, which is a different concept.

It is difficult for an external reviewer to answer the question of whether the estimates presented are correct in any general sense: neither the data nor the methods used are well enough described. Three separate topics can be identified: the distribution of the coastal stock and the corresponding offshore survey limit; the number of dolphins within the distribution area so defined; and the size of the by-catch.

<u>Distribution of the coastal stock</u>: a larger limit than 27 km should probably be taken in winter. The evidence in favour of the choice of 27 km is not convincing, and it may induce an underestimation of the stock size.

<u>Number of dolphins in the survey area</u>: certainly an underestimate, but without knowing more about turbidity, the diving behaviour of the species, and the instructions given to the observers, it is difficult to be sure by how much.

<u>By-catch</u>: the sampling design of observer coverage is not stated. The estimate for the total by-catch over the whole coastline is based on 18 observations comprising 20 animals. This is unlikely to give a credible estimate for the total, let alone for individual fisheries. There are obvious, stated, problems with effort measures for some fisheries, and there may be less obvious, unstated, problems with effort measures in others. The total by-catch is probably underestimated; again it is difficult to estimate by how much.

These studies and their results are used to set a permissible limit for by-catch via a standard expression for 'PBR'. The expression PBR—Potential *Biological* Removals—is misleading. These limits are as strongly affected by the quality of the estimation of population size as by the biology of the species: in fact the only biological factor, R_{max}, is pulled off the shelf as a blanket value for 'dolphins'. There are several problems associated with these calculations. No uncertainty is associated with R_{max}, although the value of 4%/yr given for dolphins seems low (it is thought to be 3–4%/yr even for belugas, which bear no more than once every 3 years). Secondly, the relationship between R_{max} and the sustainable by-catch depends on whether by-catch is additive to, or compensatory for, the mortality factors assumed in estimating the 4% R_{max}; this question is not discussed in these documents. Calculated 'PBR's are given as single values, with no estimate of uncertainty attached. It would be useful if the uncertainties first of the estimate of numbers, then of its CV, then of the assumed 4% R_{max}, then of the decision rule associated with F_r, were to be used to qualify the bald statement that no more than *n* dolphins could be taken from a given stock, and to generate a CV for *n*? As information gets worse, the calculated 'PBR' gets smaller, but its uncertainty increases greatly, and small 'PBRs' probably have very large ECVs. Estimated by-catches also add their own uncertainties to the equation.

Dividing the coastal population into management units will affect the calculation of PBRs. The numerical sizes of smaller stock units are estimated from smaller numbers of observations than the total coastal population, and therefore have larger ECVs.

Consequently, as individual stocks become better defined and more finely divided, total PBR over the whole fishery will tend to become smaller for a given state of knowledge of the overall population. Setting, and subsequently supervising, PBRs individually for small stocks could make large demands on quantitative stock-assessment science and on social techniques of fisheries management.

<u>Reviews of individual documents</u>

NMFS SEFSC: Preliminary stock structure of coastal bottlenose dolphins along the Atlantic coast of the U.S.

There is a problem with the selection of an offshore limit for the coastal type. The 27-km limit is not well justified. The first two sentences of section **1.2.2** may sometimes be true, but spatial patterns can exist for lots of possible reasons, and stocks may equally well overlap, at least in some seasons, with little evidence of spatial boundaries. Fig. 4 does not, to my eyes, display the 'significant gap in spatial distribution at a distance of 27 km from shore' mentioned in this section.

It seems that biopsies taken on summer cruises are being used, although important aerial surveys in the mixed-stock area take place in winter. This looks like a problem. There remains a great deal of uncertainty on the appropriate limit for aerial surveys.

Garrison: Seeking a hiatus in sightings for bottlenose dolphin during summer and winter aerial surveys.

The overall approach is original, but cumbersome: the problem seems adapted to a single-sample Kolmogorov-Smirnov test. The bootstrap approach is however reasonable, but instead of calculating all these separate confidence intervals, why not just calculate a single Kolmogorov-Smirnov-type statistic (maximum deviation between distribution functions) to measure the deviation of each synthesised distribution from the null hypothesis, obtain the distribution of this statistic, and compare the observed value of the statistic with its bootstrap distribution to obtain a single test probability?

The significance tests assume independence of sightings. To the extent that sightings are not independent, even allowing for the effect of depth/distance, this may give spurious significance levels. Are sightings independent?

Group size seems to be ignored. There may be independent data that group size does not change with distance from shore, but if so, you might refer to it somewhere.

In general, the approach you use seems to be directed at the wrong problem, i.e. detecting a drop in sightings rate. As I understand it, the problem is more to get an unbiased estimate of the size of the coastal stock. The approach which seems obvious is to consider first the proportion of dolphins that are of coastal type, *p(d)*, related to distance from shore *(d)*, and second the total number per unit distance *n(d)*. The true size of the coastal population is then:

$$N = \int_0^D n(d)\,p(d)\,dd$$

where *D* is some distance beyond which it is known that *p* is zero for all *d*, and the estimated size of the population is

$$\hat{N} = \int_0^L n(d)dd$$

with *L*, the survey limit or estimation limit, less than *D*. The bias in estimation is then

$$B = \hat{N} - N = \int_0^L n(d)(1 - p(d))dd - \int_L^D n(d)p(d)dd \, .$$

The first of these terms represents offshore dolphins reckoned in the coastal population and increases with L; the second represents coastal dolphins not so counted and decreases with increasing L. The objective is to find L so that the bias is zero. I don't suppose that the data available will support an exact solution to this equation. But at least this formulation allows the error in stock size estimation to be described, and also allows quantifying the uncertainty in terms of the uncertainties in the available data. It may allow calculating some kind of component of error in the estimate of the coastal stock size(s). It seems as though 39 km would be an initial estimate for D, and that L is greater than 6 km, at least in summer.

Garrison and Hohn. Abundance estimates for Atlantic bottlenose dolphin: combining strip transect data and line transect abundance estimation.

Confusion ensues from referring to these coastal surveys as 'strip transect' surveys. A transect is (L) a 'cut across' the habitat or study area, by implication and usage straight; not quite what we have here. It would be helpful to call these simply 'coastal strip surveys'.

The method used to obtain the ECV is unclear. The only way you describe the method of estimating CVs is 'We employed a similar method to estimate CV' (it would be helpful if you said similar to what) and 'the ratio of encounter-rate CVs was used' (to do what?). The standard basic expression for ECVs of ratios or products (of uncorrelated terms) is the square root of the sum of squares of the ECVs of the terms, and it can be found in any elementary statistics text. It is not clear from your text that this is what you used; I suspect that you simply guessed at an expression and came up with a wrong guess. Note that the ratio of CVs *obviously* cannot be a valid expression for the CV of a ratio, because when the ratio is expressed the other way up, its CV remains about the same, whereas the ratio of the CVs will also be inverted.

The intention is to extrapolate the line-transect surveys to unsurveyed areas using the coastal strip survey observations. The source of uncertainty in doing so is the spatial variability in the ratio between the coastal survey observations and the line-transect observations. If this ratio did not vary much from place to place, one would expect to be able to extrapolate with confidence. If it did so vary, one would not put much confidence in one's extrapolation. But the temporal variability in the strip survey counts has little part to play. If the spatial variability were high, the extrapolation would be unreliable even if the temporal variability were zero.

A possible approach would be to tabulate, transect by transect, the line-transect density estimate against the strip-survey counts along the associated segment of strip. The result would be a two-way table of strip-survey counts with dates as rows and transects with columns, with one additional row of line-transect density estimates. The line-transect density estimates would be missing in the extrapolation area. The objective of the analysis would be to construct a suitable model of the relationship between the strip-survey counts in the area covered by the line-transect survey, and to use that relationship to fill in the sum of the missing line-transect densities in the extrapolation area, or their sum. The analysis might be complex; however, in this representation, the variation between individual coastal strip surveys is represented by the between-rows mean square, and that statistic does not appear to have much to do with the uncertainty attaching to the extrapolated densities.

Extrapolation is always dangerous. If the spatial variability of the encounter-rate ratios shows serial correlation between transects, (i.e. there are local trends in the encounter-rate ratios), it becomes possible that some such trend in the true ratio within the extrapolation area could lead an estimate that was calculated ignoring such a possibility much astray. An analysis of, and a statement on, the between-transect spatial-series properties of the encounter-rate ratios would help the extrapolation to be more acceptable.

Garrison and Yeung: Abundance estimates for Atlantic bottlenose dolphin stocks during summer and winter 1995.

Page 3 is a little loose, as far as precise description of the methods used is concerned. In Equation 2, the limit of integration *w* (said to be the esw) is not defined; even if it were, the equation is wrong. The (uncorrected) survey expansion factor (your *f*(0)) is the reciprocal of the esw, and the esw is the integral of the sighting curve (corrected so g_{max} is unity) from zero to infinity. Although the calculation of confidence limits from the error variance of the estimate is described, the method of calculating the error variance itself is not.

Methods used for treating sighting data might well be revised. You:

a) Started off with an assumption of monotonically decreasing visibility with increasing distance away. Anyone who has looked out of the side window of an aircraft would probably question this assumption;

b) Having found that the data didn't agree with the untenable assumption, discarded one-fifth to one-quarter of the data so that it did agree. Better scientific practice would be to change the assumption, not the data;

Further than that the methods are poorly described. It is likely that model-selection error was ignored, so that all ECVs are biased downwards, but as you don't say how ECVs were calculated, it is difficult to assess this.

What is meant by 'subjectively grouped into several different interval categories'? Did you try several different ways of histogramming the data until you found one you liked? If so, did you estimate a component of model-selection error to account for this? It is recommended to fit two-sided sighting curves that will model the increase of visibility with distance that is commonly experienced close to side-looking platforms, and to fit them to the data as recorded, without histogramming it.

For the summer survey, a one-sided sighting curve was then fitted to the data, and the density estimates reported are, the peak densities that were really seen. The ECV of the esw presumably includes no component of uncertainty for the choice of model, or the choice of histogramming, but these may be small. The population estimates are based on mean densities in a wide strip of uniform visibility between 100 m and about 250 m. The reported ECV of 9% looks too small, but not much too small, given the data. (Note that the density estimate for 'Northern Migratory' appears to be based on an esw of 247.9 m rather than the 244.9 m reported.)

For the winter survey the peak sighting densities recorded, i.e. those actually seen out of the window of the aircraft at 100 m distance, have been fudged upwards, by a very doubtful amount. Fitting a sighting curve for line-transect analysis when data exists all the way to the track-line is undemanding as to the model selected, so for most regular line-transect work a small choice of sighting models suffices. But when it is a question of extrapolating back across a 100-m gap, the sensitivity to model selection is of a very different order, and it is likely that many different models could fit the data about as well as each other but would yield very different multipliers. Although the increase used here is only about 10%, there must be a large uncertainty of model selection associated with it; it would help if you could account for it in your stated ECV.

This is a big difference in treatment for the two surveys: for the summer survey, estimates are based on mean densities between 100 m and 250 m; for the winter surveys, even observations at 100 m are considered inadequate; and this with the same platform flying at the same height and speed, under similar conditions, and targeting the same species. There is no discussion of this very different treatment; there should be; it may be due to observers receiving different instructions in the two surveys. For example, if the winter-survey observers were told to look *down*, and therefore failed to see *out*, their peak densities close to the aircraft could have been increased, and the further upward alteration would increase an existing bias. The results of these two surveys are not comparable.

Heuristically, the logic is that the summer observers could see as much at 250 m as at 100 m, and this is interpreted as indicating that they were so good that their 100-m counts must be perfect. The winter observers at 250 m could only see 61% of what they saw at 100, so they were so bad that even their 100-m sightings had to be corrected. However, another logic is possible: the summer observers could only see at 100 m as much as they saw at 250, so their 100-m counts couldn't be very good and ought to be adjusted, but the winter observers saw at 100 m fully 63½% more than at 250 m, so their 100-m counts had to be pretty good. Why is one logic better than the other? I do have a lot of doubts about increasing the 100-m density observed in the winter survey simply because it is better than the 250-m density.

The use of two-sided sighting curves capable of fitting the increase in visibility with distance that is common when side-looking platforms are used has much to recommend it. One of the benefits to be looked for from line-transect surveys is to transfer information, in the form of the fitted sighting curve, from densely populated areas for use in improving estimates in sparsely populated ones. If you are having to throw out 23% of the observations, some of this benefit is being lost.

The methods used to estimate the total number, and the component of error variance due to the uncertainty in encounter rate, are not stated. The summer surveys seem to have a fairly regular systematic design, and the appropriate estimate of encounter-rate uncertainty would be by serial difference. The winter surveys have an odd design, with a small number of blocks, each systematically sampled but not all at the same spacing, separated by gaps. Were transects weighted by the spacing in their blocks; if not, why not? If there is serial correlation between encounter rates, this sampling, with gaps between blocks, may be less precise than random sampling (i.e. the precision gained by systematic sampling within blocks may be more than lost by the gaps between blocks). The serial correlation between transects should be analysed. Might the strip-survey analysis that was used to extrapolate the line-transect density estimate outside the line-transect survey area also be used to improve the interpolation within these gaps in the line-transect design?

Palka, Garrison, Hohn and Yeung: Summary of abundance estimates and PBR for coastal *Tursiops* for waters between New York and Florida during 1995 to 2000.

I suggest adding, to the description of each survey, information on its design; transect spacing, stratification, &c.(What is this?), including perhaps what it was intended to measure and why it was designed the way it was. The MATS surveys look like ordinary unstratified systematic sampling; the SECAS programme looks like 'stratified systematic with gaps'. Why the stratification? Why the gaps? Sampling design of the NE programme is not stated.

Many of these estimates are correlated. The strip-survey extensions of the line-transect estimates will be strongly correlated with the line-transect estimates. This probably deserves to be mentioned.

Palka and Rossman: Bycatch estimates of coastal bottlenose dolphin (*Tursiops truncatus*) in US mid-Atlantic gillnet fisheries for 1996 to 2000.

The GLM model of by-catch factors is interesting, and the factors that remain in the best model appear to make sense. Given that anchor use is associated with large-mesh (if I correctly understand your text), and that bringing in anchor use improves the model, I wondered if anchor use coming *in* tends to drive large mesh *out* of the model's parameter set.

A point of concern in this document is the estimated ECV of the by-catch. When observations consist of independent rare events, the ECV of consequent estimates of rates or totals should be close to the reciprocal of the square root of the number of events observed (from the mean-variance relationship of the Poisson distribution). It appears that the present estimates are based on 11 observed dead dolphins; you don't mention any larger number and the highlighted dead takes in Table 3 sum to 11. The reciprocal of the sq. root of 11 is about 30%, so the basic ECV of any resulting estimate of total by-catch *starts* at around 30%. (Note in passing that Garrison, calculating the by-catch in the gillnet fishery for sharks, obtains ECV values that *are* close to the predicted $1/\sqrt{n}$.)

This value of about 30% is the starting point. Variation in observer coverage between fisheries will increase it (See App. 1). So will any uncertainty in observer coverage, or landings (and some uncertainty might be expected in the factors used to get us from 11 dolphins to 233? For example, landings are allocated to mesh size class on the basis of target species, which is not a certain translation.) So when your estimated ECV of *half* this value comes along with no explanation or discussion, it's not easy to believe it; where can the extra information have come from to get such a small ECV from so few observations? Since it seems that the overall estimate of the annual total by-catch was obtained by summing the by-catch from the individual fisheries, similar considerations presumably apply to individual years' values.

In calculating averages over time, you say that you weighted annual values according to their precision. This common approach to combining values may be acceptable when precision truly reflects effort (as when samples of different sizes are drawn from the same population). When (as in values resulting from observations of rare events) the precision mostly reflects a (random) number of events observed, this rule is likely to result in biased estimates of the mean and of its error variance (see App. II) and weighting by effort directly is often more appropriate. It would be helpful if you at least justified this decision; but without more information on the fishery it is not easy to know how to combine annual estimates to estimate the long-term mean and its uncertainty.

P. 21 mentions low observer coverage ('too low in most years to have seen a take') as a source of bias in by-catch estimation ('by-catch estimate is biased low because of low coverage of a small fishery'). This is a fallacy: low coverage increases uncertainty, but estimates of an observed rate are not biased if effort is applied randomly with respect to events.

Landings are a poor measure of effort, but there seems to be no obvious alternative; the procedure relies on recording being complete. Can observer data be checked against the landings database to check whether landings are complete? Are there no logbooks in these fisheries?

Garrison: Mortality estimate for Atlantic bottlenose dolphin in the directed shark gillnet fishery off Florida and Georgia.

This seems clear and straightforward. The stratification is reasonable. The observations on missing logbook data might be used to make a quantitative correction to the bycatches? That significant gillnet effort is directed towards sharks outside the documented fleet makes this exercise look irrelevant unless that effort is measured and included in the uncertainty.

All in all, the effort data in these fisheries seems so unreliable that these bycatch estimates may not be useful, except as an indicator that bycatch exists and may be significant.

Appendices

Appendix I: Effect of variation in observer coverage on ECV of estimates of total by-catch.

Fishery 1 has n_1 by-caught dolphins and observer coverage p_1. Its resulting estimate of total by-catch is n_1/p_1 with ECV $1/\sqrt{n_1}$ and error variance n_1/p_1^2. Fishery 2 has n_2 by-caught dolphins and observer coverage p_2. Its resulting estimate of total by-catch is n_2/p_2 with ECV $1/\sqrt{n_2}$ and error variance n_2/p_2^2. The estimated total by-catch is $n_1/p_1 + n_2/p_2$ with variance $n_1/p_1^2 + n_2/p_2^2$ and therefore ECV $\sqrt{(n_1/p_1^2 + n_2/p_2^2)} \div (n_1/p_1 + n_2/p_2)$. If p_2 is $p_1\alpha$, this can be written $\sqrt{(n_1\alpha^2 + n_2)} \div (n_1\alpha + n_2)$. This is minimised when $\alpha=1$, and is then $1/\sqrt{(n_1+n_2)}$, which is what it should be. As α becomes large, this expression tends to $1/\sqrt{n_1}$, and when α becomes small, it tends to $1/\sqrt{n_2}$.

Appendix II: Average values of rates.

Example: a process in which events are random and independent is observed for an hour, and 1 event is observed; the estimate of the mean is 1 event/hr with error variance 1 (event/hr)^2. The same process is then observed for two more hours and 3 events are observed, giving another estimate of the mean of 1.5/hour with error variance ¾ (ev/hr)^2. If an average is calculated by conventionally weighting the two estimates according to their error variances, the result is a weighted mean of 9/7 events/hour with error variance 3/7. The correct values (from combining the two observation periods) are obviously 4/3 events/hour for the mean and 4/9 for the variance. The incorrect weighting biases the mean upwards and the variance downwards.

Appendix III: Model-selection error.

Properties of statistical estimates of parameters, including their precision, are usually conditional on the statistical model fitted being an accurate representation of the process that generated the data. If this is the case, the properties of the parameter estimates can be accurately described. However, the process is often unknown. In such cases, the selection of a model is an important step in a statistical analysis. Statistical orthodoxy assumes that a model fitted to data is identified by means external to the data. If a single data set is used both to select a model and to estimate parameter values for the selected model, the precision of the resulting estimates is overestimated and predictions made using the model will be wrong unexpectedly often.

This statistical problem is fairly well recognised in the form of 'posterior stratification'. It occurs in many other contexts. Stepwise fitting of regression and other models typically overestimates significance levels unless corrections are applied. In line-transect analysis, using the same data set both to select a visibility curve and to estimate its parameters can be expected to underestimate the uncertainty of the esw, the problem being compounded if the visibility-curve models are fitted to a heuristically-selected histogramming of the data. Line-transect analysis may in fact be regarded as a stratified analysis of count data, selecting a visibility model on the basis of the data being equivalent to post-stratification.

Prior selection of a model is preferable: either through prior knowledge of the process, by pilot study, or by dividing the data and using one part to select a model, and the other part to estimate parameter values. It might however be possible to use a resampling approach, repeating the model-selection step at each bootstrap iteration. This would perhaps allow reasonable estimation of the uncertainty attaching to derived parameters such as, in line-transect analysis, the esw.

Appendix IV: Statement of Work

STATEMENT OF WORK

CONSULTING AGREEMENT BETWEEN THE UNIVERSITY OF MIAMI AND DR. MICHAEL KINGSLEY

marts 12, 2002

GENERAL[1]

THE MARINE MAMMAL PROTECTION ACT REQUIRES THE NATIONAL MARINE FISHERIES SERVICE (NMFS) TO DEVELOP AND IMPLEMENT A TAKE REDUCTION PLAN TO ASSIST IN THE RECOVERY OR PREVENT THE DEPLETION OF STRATEGIC STOCKS OF MARINE MAMMALS THAT INTERACT WITH COMMERCIAL FISHERIES THAT FREQUENTLY (CATEGORY I) OR OCCASIONALLY (CATEGORY II) CAUSE INCIDENTAL MORTALITY OR SERIOUS INJURY TO MARINE MAMMALS. TAKE REDUCTION TEAMS ARE CONVENED TO DEVELOP DRAFT TAKE REDUCTION PLANS, WHICH ARE IMPLEMENTED BY NMFS THROUGH REGULATIONS. TAKE REDUCTION TEAMS CONSIST OF REPRESENTATIVES FROM THE COMMERCIAL AND RECREATIONAL FISHING INDUSTRY, CONSERVATION GROUPS, FEDERAL AND STATE GOVERNMENT, FISHERY MANAGEMENT COUNCILS, INTERSTATE FISHERIES COMMISSIONS, AND ACADEMIC AND SCIENTIFIC ORGANIZATIONS.

THE IMMEDIATE GOAL OF A TAKE REDUCTION PLAN IS TO REDUCE, WITHIN 6 MONTHS OF IMPLEMENTATION, THE INCIDENTAL MORTALITY OR SERIOUS INJURY OF A MARINE MAMMAL STOCK FROM COMMERCIAL FISHING OPERATIONS TO A SUSTAINABLE LEVEL, REFERRED TO AS THE POTENTIAL BIOLOGICAL REMOVAL LEVEL. THE TAKE REDUCTION PROCESS IS CONTENTIOUS, BRINGING TOGETHER PEOPLE WITH VERY DIFFERENT PERSPECTIVES TO DEVELOP A CONSENSUS-BASED APPROACH FOR REDUCING MARINE MAMMAL MORTALITY INCIDENTAL TO COMMERCIAL FISHING. THE INTENT IS TO DEVELOP A MANAGEMENT PROGRAM THAT MEETS CONSERVATION GOALS AND MINIMIZES THE POTENTIAL IMPACT ON THE FISHING INDUSTRY.

THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM (TEAM) WAS CONVENED IN NOVEMBER OF 2001 TO ADDRESS MORTALITY OF THE WESTERN NORTH ATLANTIC COASTAL STOCK OF BOTTLENOSE DOLPHINS INCIDENTAL TO NINE CATEGORY II COMMERCIAL FISHERIES THAT OCCUR ALONG THE EAST COAST OF THE UNITED STATES. THE WESTERN NORTH ATLANTIC COASTAL STOCK OF BOTTLENOSE DOLPHINS IS A STRATEGIC STOCK. STRATEGIC STATUS WAS

---

[1] See Attachment A for definitions of the following terms: strategic stock, potential biological removal level, depleted, optimum sustainable population, and fishery classifications (Category I, II, and III fisheries).

INITIALLY ASSIGNED BECAUSE THE STOCK IS DESIGNATED AS DEPLETED UNDER THE MARINE MAMMAL PROTECTION ACT AS A RESULT OF A LARGE-SCALE MORTALITY EVENT THAT OCCURRED IN 1987-1988. HOWEVER, THE STOCK ALSO QUALIFIES TO BE STRATEGIC BECAUSE MORTALITY AND SERIOUS INJURY INCIDENTAL TO COMMERCIAL FISHING EXCEEDS SUSTAINABLE LEVELS.

THE DATA USED IN THE ANALYSES CONDUCTED TO DEVELOP STOCK STRUCTURE, ABUNDANCE, AND FISHERY-RELATED MORTALITY INFORMATION FOR USE BY THE TEAM ARE NOT FINAL AND WILL CONTINUE TO BE SUPPLEMENTED BY RESULTS FROM ONGOING RESEARCH EFFORTS. THE DOCUMENTS SUBMITTED FOR REVIEW HAVE ALREADY BEEN PEER-REVIEWED BY THE ATLANTIC SCIENTIFIC REVIEW GROUP THROUGH A SYSTEM ESTABLISHED BY THE MARINE MAMMAL PROTECTION ACT. MEMBERS OF THE SCIENTIFIC REVIEW GROUPS ARE INDIVIDUALS WITH EXPERTISE IN MARINE MAMMAL BIOLOGY AND ECOLOGY, POPULATION DYNAMICS AND MODELING, AND COMMERCIAL FISHING TECHNOLOGY PRACTICES.

AS REQUIRED BY THE MARINE MAMMAL PROTECTION ACT, NMFS HAS CONVENED THE TEAM AND MUST USE THE BEST AVAILABLE INFORMATION TO SUPPORT THE TEAM. ALTHOUGH THE TEAM IS REQUIRED BY THE MARINE MAMMAL PROTECTION ACT TO SUBMIT A DRAFT TAKE REDUCTION PLAN TO NMFS IN MAY OF 2002, THEIR INVOLVEMENT IN THE PROCESS DOES NOT END AT THAT TIME. TAKE REDUCTION TEAMS CONTINUE TO MEET WITH NMFS ON A REGULAR BASIS TO MONITOR PLAN IMPLEMENTATION AND RECEIVE NEW INFORMATION RESULTING FROM ONGOING RESEARCH EFFORTS. AT THESE MEETINGS, THE TEAM HAS THE OPPORTUNITY TO MAKE RECOMMENDATIONS TO REVISE THE REGULATIONS. ATTACHMENT B SHOWS THE CHRONOLOGY OF EVENTS RELATED TO CONVENING THE TEAM.

SPECIFIC

EACH REVIEWER SHALL ANALYZE SEVEN DOCUMENTS ADDRESSING BOTTLENOSE DOLPHIN STOCK STRUCTURE, ABUNDANCE ESTIMATES, AND MORTALITY ESTIMATES (TASK 2, ITEMS A-G). THE SEVEN DOCUMENTS REPRESENT AN INTERIM APPROACH TO BOTTLENOSE DOLPHIN SCIENCE. NMFS RECOGNIZES THAT THEY PRESENT WORK IN PROGRESS (E.G., STOCK IDENTIFICATION), *A POSTERIORI* ANALYSIS OF PREVIOUS DATA SETS USING NEW STOCK STRATA (E.G., SOUTHEAST SURVEYS FOR ABUNDANCE), AND, IN SOME CASES, LIMITED SAMPLE SIZES. HOWEVER, THIS INFORMATION IS A SIGNIFICANT IMPROVEMENT OVER THE INFORMATION AVAILABLE PRIOR TO THE COMPILATION OF THESE DOCUMENTS AND, THEREFORE, PROVIDES MORE APPROPRIATE RESULTS FOR USE BY THE TEAM. EACH REVIEWER SHALL REVIEW THESE DOCUMENTS IN THAT CONTEXT.

THE REVIEWERS SHOULD FOCUS ON THE FOLLOWING ISSUES WHEN WORKING ON TASK 2:

- ARE THE DATA USED APPROPRIATELY?
- ARE THE ANALYSES CONDUCTED APPROPRIATE?
- ARE THE DATA ADEQUATE FOR THE ANALYSES CONDUCTED?

- Are assumptions inherent in the analyses appropriately considered?
- Are the interpretations of the data and analyses appropriate?

Additionally, the reviewers shall provide suggestions for alternative methods for analyzing or interpreting the information, where better methods exist. Each reviewer shall conclude, in a written report, whether the analyses represent the best analysis of the available information.

Each reviewer's duties shall not exceed a maximum total of seven days, including several days for document review, several days to produce a written report of the findings, and several hours for a conference call. No travel is required, and therefore each reviewer may perform all review, analysis, and writing duties out of the reviewer's primary location. A consensus report is not required.

The Center of Independent Experts (CIE) shall schedule and facilitate a conference call with the reviewers and NMFS scientists and managers to provide an opportunity for the reviewers to ask questions about the science and the take reduction process prior to the reviewers writing their reports as detailed under Task 4 and Annex I. The agenda for the conference call shall consist of: 1) general/programmatic issues; (2) questions specific to each document listed in Task 2. The CIE shall provide a toll-free phone number for the conference call. Ideally, one conference call will be held with all of the reviewers. However, more than one conference call may be arranged if all of the reviewers are not available at the same time.

The itemized tasks of each reviewer include:

<u>Task 1</u>: Surveying the following supplementary information provided to each reviewer prior to completing Task 2 for additional context and background about bottlenose dolphins, the take reduction process, and the Atlantic Marine Mammal Scientific Review Group review. The reviewer should not analyze these documents for the written report.

The following two documents provide an overview of bottlenose dolphin science, as presented in NMFS Stock Assessment Report publications. The first document represents the current state of knowledge, summarizing the detailed information contained in the review documents identified in Task 2. The second document is provided to show what information was available before the review documents were prepared.

- National Marine Fisheries Service. November 2001. Draft 2002 Stock Assessment Report for the Western North Atlantic Coastal Stock of Bottlenose Dolphin (*Tursiops truncatus*).

- National Marine Fisheries Service. September 2000. 2000 Stock Assessment Report for the Western North Atlantic Coastal Stock of Bottlenose Dolphin (*Tursiops truncatus*).

The following two documents relate to the review conducted by the Atlantic Scientific Review Group, which reviewed the same or earlier versions of the documents identified in Task 2.

- Atlantic Scientific Review Group review of Bottlenose Dolphin documents. October 2001.

- NMFS response to the Atlantic Scientific Review Group. November 2001.

NMFS gave Team members the opportunity to identify questions about the documents identified in Task 2. Only one Team member provided comments. The reviewers are not required to respond to these comments.

- Comments from one member of the Team for the CIE peer review. December 2001.

The following two documents identify concerns of one representative from the fishing industry about bottlenose dolphin science and the take reduction process and NMFS response.

- Letter from Rick Marks to the Honorable James V. Hansen and the Honorable Don Young of the U.S. House of Representatives Resources Committee regarding the bottlenose dolphin take reduction team process. August 2001.

- NMFS response to Rick Marks letter to the House Resources Committee. September 2001.

The following three documents provide descriptive information about bottlenose dolphin strandings along the east coast of the U.S.

- Palka, D., F. Wenzel, D. L. Hartley, and M. Rossman. June 2001. Summary of bottlenose dolphin strandings from New York to Virginia.

- Hohn A., P. T. Martone. July 2001. Characterization of bottlenose dolphin strandings in North Carolina, 1997-2000.

- Hohn A., B. Mase, J. Litz, W. McFee, and B. Zoodsma. November 2001. Characterization of human-caused strandings of bottlenose dolphins along the Atlantic coast from South Carolina to Florida, 1997-2000.

TASK 2: READING AND ANALYZING THE FOLLOWING DOCUMENTS (A-G) PROVIDED TO EACH REVIEWER. THIS IS THE PRIMARY TASK OF THE CONTRACT. THE REPORT IDENTIFIED IN TASK 4 AND IN ANNEX I SHOULD ADDRESS THESE DOCUMENTS.

STOCK STRUCTURE

a. NATIONAL MARINE FISHERIES SERVICE. JUNE 2001. PRELIMINARY STOCK STRUCTURE OF COASTAL BOTTLENOSE DOLPHINS ALONG THE ATLANTIC COAST OF THE U.S.
b. GARRISON, L. JUNE 2001. SEEKING A HIATUS IN SIGHTINGS FOR BOTTLENOSE DOLPHIN DURING SUMMER AND WINTER AERIAL SURVEYS. NATIONAL MARINE FISHERIES SERVICE.

ABUNDANCE ESTIMATES

C. GARRISON, L. AND A. HOHN. OCTOBER 2001. ABUNDANCE ESTIMATES FOR ATLANTIC BOTTLENOSE DOLPHINS: COMBINING STRIP TRANSECT DATA AND LINE TRANSECT ABUNDANCE ESTIMATION. NATIONAL MARINE FISHERIES SERVICE.
D. GARRISON, L. AND C. YEUNG. 15 JUNE 2001. ABUNDANCE ESTIMATES FOR ATLANTIC BOTTLENOSE DOLPHIN STOCKS DURING SUMMER AND WINTER, 1995. NATIONAL MARINE FISHERIES SERVICE.

E. PALKA, D., L. GARRISON, A. HOHN, AND C. YEUNG. 1 NOVEMBER 2001. SUMMARY OF ABUNDANCE ESTIMATES AND PBR FOR COASTAL *TURCIOPS* FOR WATERS BETWEEN NEW YORK AND FLORIDA DURING 1995 TO 2000. NATIONAL MARINE FISHERIES SERVICE.*MORTALITY ESTIMATES*
F. GARRISON, L. 2 JULY 2001. MORTALITY ESTIMATE FOR ATLANTIC BOTTLENOSE DOLPHIN IN THE DIRECTED SHARK GILLNET FISHERY OF FLORIDA AND GEORGIA. NATIONAL MARINE FISHERIES SERVICE.
G. ROSSMAN, M. AND D. PALKA. 3 OCTOBER 2001. BYCATCH ESTIMATES OF COASTAL BOTTLENOSE DOLPHIN (*TURCIOPS TRUNCATUS*) IN U.S. MID-ATLANTIC GILLNET FISHERIES FOR 1996 TO 2000. NATIONAL MARINE FISHERIES SERVICE.

TASK 3: PARTICIPATE IN A CONFERENCE CALL, TO BE ARRANGED BY CIE, WITH NMFS SCIENTISTS AND MANAGERS TO DISCUSS QUESTIONS EACH REVIEWER MAY HAVE ABOUT THE SCIENCE AND THE TAKE REDUCTION PROCESS.

TASK 4: NO LATER THAN MARCH 1, 2002, EACH REVIEWER SHALL SUBMIT A WRITTEN, NON-CONSENSUS REPORT OF FINDINGS, ANALYSIS, AND CONCLUSIONS BASED OF THEIR REVIEW OF THE DOCUMENTS (TASK 2, ITEMS A-G). THE REPORT SHOULD BE ADDRESSED TO THE UNIVERSITY OF MIAMI INDEPENDENT SYSTEM FOR PEER REVIEWS AND SENT TO DR. DAVID

DIE, UNIVERSITY OF MIAMI/RSMAS, 4600 RICKENBACKER CAUSEWAY, MIAMI, FL 33149 (OR VIA EMAIL TO DDIE@RSMAS.MIAMI.EDU).

SIGNED________________ DATE_______________

ANNEX I: REPORT GENERATION AND PROCEDURAL ITEMS

3. THE REPORT SHOULD BE PREFACED WITH AN EXECUTIVE SUMMARY OF FINDINGS AND/OR RECOMMENDATIONS.

4. THE MAIN BODY OF THE REPORT SHOULD CONSIST OF A BACKGROUND, DESCRIPTION OF REVIEW ACTIVITIES, SUMMARY OF FINDINGS, AND CONCLUSIONS/RECOMMENDATIONS.

5. THE REPORT SHOULD ALSO INCLUDE AS SEPARATE APPENDICES THE BIBLIOGRAPHY OF MATERIALS PROVIDED BY THE CENTER FOR INDEPENDENT EXPERTS AND NMFS AND A COPY OF THE STATEMENT OF WORK.

Attachment A
Definitions

Strategic stock is defined in section 3(19) of the MMPA to mean a marine mammal stock, "(A) for which the level of direct human-caused mortality exceeds the potential biological removal level; (B) which, based on the best available scientific information, is declining and is likely to be listed as a threatened species under the Endangered Species Act (ESA) of 1973 within the foreseeable future or (C) which is listed as a threatened species or endangered species under the Endangered Species Act of 1973 (16 U.S.C. 1531 et seq.), or is designated as depleted under this Act [MMPA]."

Potential biological removal (PBR) level is defined in section 3(20) of the MMPA to mean, in relevant part, "the maximum number of animals, not including natural mortalities, that may be removed from a marine mammal stock while allowing that stock to reach or maintain its optimum sustainable population."

Depleted is defined by section 3(1) of the MMPA to mean any case in which, "(A) the Secretary, after consultation with the Marine Mammal Commission and the Committee of Scientific Advisors on Marine Mammals established under Title II of this Act, determined that a species or population stock is below its optimum sustainable population; (B) a State, to which authority for the conservation and management of a species or population stock is transferred under section 109, determines that such species or population stock is below its optimum sustainable population; or (C) a species or population stock is listed as an endangered species or a threatened species under the Endangered Species Act of 1973."

Optimum sustainable population is defined by section 3(9) of the MMPA to mean, "with respect to any population stock, the number of animals which will result in the maximum productivity of the population or the species, keeping in mind the carrying capacity of the habitat and the health of the ecosystem of which they form a constituent element."

Fishery classification is defined in section 118(c) of the MMPA and implemented by regulation in 50 CFR part 229. The fishery classification criteria consist of a two-tiered, stock-specific approach that first addresses the total impact of all fisheries on each marine mammal stock and then addresses the impact of individual fisheries on each stock. This approach is based on consideration of the rate, in numbers of animals per year, of incidental mortalities and serious injuries of marine mammals due to commercial fishing operations relative to the PBR level for each marine mammal stock.

- <u>Tier 1</u>: If the total annual mortality and serious injury across all fisheries that interact with a stock is less than or equal to 10 percent of the PBR level of this stock, all fisheries interacting with this stock would be placed in Category III. Otherwise, these fisheries are subject to the next tier of analysis to determine their classification.

- <u>Tier 2, Category I</u>: Annual mortality and serious injury of a stock in a given fishery is greater than or equal to 50 percent of the PBR level.
- <u>Tier 2, Category II</u>: Annual mortality and serious injury of a stock in a given fishery is greater than 1 percent and less than 50 percent of the PBR level.
- <u>Tier 2, Category III</u>: Annual mortality and serious injury of a stock in a given fishery is less than or equal to 1 percent of the PBR level.

ATTACHMENT B
CHRONOLOGY OF EVENTS RELATED TO THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM

1997 NMFS INITIATED TAKE REDUCTION EFFORTS FOR BOTTLENOSE DOLPHINS IN 1997 AS PART OF THE MID-ATLANTIC TAKE REDUCTION TEAM. AT THE TIME, DATA ON BOTTLENOSE DOLPHIN STOCK STRUCTURE, ABUNDANCE, AND MORTALITY WERE LIMITED. IN 1997, THE LACK OF DATA MADE DISCUSSION OF SOLUTIONS TO REDUCE BYCATCH DIFFICULT, AND THEREFORE NMFS DELAYED ADDRESSING BOTTLENOSE DOLPHIN ISSUES UNTIL BETTER DATA WERE AVAILABLE.

NOV. 2000 NMFS PRESENTED INITIAL INFORMATION ON BOTTLENOSE DOLPHIN SCIENCE AT THE ANNUAL MEETING OF THE ATLANTIC MARINE MAMMAL SCIENTIFIC REVIEW GROUP AND RECEIVED RECOMMENDATIONS FOR MODIFYING THE ANALYSIS.

APRIL 2001 INTERIM REVIEW OF BOTTLENOSE DOLPHIN SCIENCE BY THE ATLANTIC MARINE MAMMAL SCIENTIFIC REVIEW GROUP.

MAY 2001 NMFS SPONSORED THE FIRST OF TWO WORKSHOPS WITH THE PUBLIC TO SHARE DATA AND OTHER INFORMATION RELATIVE TO BOTTLENOSE DOLPHIN INTERACTIONS WITH COMMERCIAL FISHERIES. PRESENTATIONS FOCUSED ON STOCK STRUCTURE, ABUNDANCE, AND DISTRIBUTION OF BOTTLENOSE DOLPHINS.

JULY 2001 NMFS SPONSORED THE SECOND OF TWO WORKSHOPS WITH THE PUBLIC TO SHARE DATA AND OTHER INFORMATION RELATIVE TO BOTTLENOSE DOLPHIN INTERACTIONS WITH COMMERCIAL FISHERIES. PRESENTATIONS FOCUSED ON THE BIOLOGY OF BOTTLENOSE DOLPHINS, MORTALITY, BOTTLENOSE DOLPHIN STRANDINGS, AND FISHERIES WITH A HISTORY OF INTERACTIONS WITH BOTTLENOSE DOLPHINS.

OCT 2001 FORMAL REVIEW OF BOTTLENOSE DOLPHIN SCIENCE BY THE ATLANTIC MARINE MAMMAL SCIENTIFIC REVIEW GROUP.

NOV 2001 FIRST MEETING OF THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM

JAN 2002 SECOND MEETING OF THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM

FEB 2002 THIRD MEETING OF THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM

MAR 2002 FOURTH MEETING OF THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM

APRIL 2002 FIFTH AND FINAL MEETINGS OF THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM

MAY 2002 FINAL REPORT WITH RECOMMENDATIONS FOR REDUCING BOTTLENOSE DOLPHIN BYCATCH DUE FROM THE BOTTLENOSE DOLPHIN TAKE REDUCTION TEAM TO NMFS.

SUMMER 2002 NMFS PUBLISHES PROPOSED REGULATIONS IMPLEMENTING THE BOTTLENOSE DOLPHIN TAKE REDUCTION PLAN AND SOLICITS PUBLIC COMMENT ON THE PROPOSED RULE.

WINTER 2003 NMFS PUBLISHES FINAL REGULATIONS IMPLEMENTING THE BOTTLENOSE DOLPHIN TAKE REDUCTION PLAN.