Quantification of Uncertainty in Fishery Stock Assessments: Statistical Challenges in the Provision of Management Advice for Southern Bluefin Tuna

Tom Polacheck
CSIRO Marine Research
GPO Box 1538
Hobart, Tasmania 7001
Tom.Polacheck@marine.csiro.au

1. Introduction

The quantitative estimation of the status of fishery resources has dominated fishery science and management. The estimation problem is generally carried out in the context of a population dynamics model and the results (referred to as stock assessments) constitute the primary scientific advice for managing fisheries. While statistics provides important tools, more statistical research is needed for stock assessment. To integrate the various relevant pieces of data from the commercial fishery, scientific surveys and other sources, better parametric models for the population dynamics of the stock and the dynamics of the fishery are required. Methods of analysis should handle the multiplicity, complexity and statistical limitations of the data. It should also lead to reliable quantification of the uncertainty and a robust statistical estimation framework to ensure that the risk of over-fishing is "very low" (United Nations, 1995).

2. The Basic Problem

The primary data available for assessing many fish stocks are time series of the total catches, the amount of fishing effort expended and samples for estimating the size and/or age composition of the catch. The spatial and temporal scales of available data will vary among fisheries, as will be the amount of relevant details on fishing activities. Focusing on the process of mortality, the following equations are fundamental.

$$N_{i+1,t+1} = N_{i,t} e^{-(F_{i,t}+M_{i,t})}, \quad C_{i,t} = \frac{F_{i,t}}{F_{i,t}+M_{i,t}} N_{i,t}(1-e^{-(F_{i,t}+M_{i,t})}).$$

Here $N_{i,t}$= the number of fish; $C_{i,t}$= the actual catch (in numbers); $F_{i,t}$ and $M_{i,t}$= fishing and natural mortality rates, and the sub-script i and t represent age and time respectively. Natural mortality rates ($M_{i,t}$) are often assumed to be constant over time and age, and are treated as known parameters. There are then $A\cdot(T+1)$ free parameters but only $A\cdot T$ unique catch observations (A is the number of age-classes and T is the number of years of observations). Additional data and/or constraints are required to identify and estimate the model.

In traditional VPA (Virtual Population Analysis), this problem of over-parameterization is solved by assuming the catch-at-age ($C_{i,t}$) to be observed without error. There are then A+T-1 free parameters. In the alternative separable VPA, the fishing mortality is assumed multiplicative in age- and year effects, $F_{i,t}= S_i\cdot G_t$. However, even in the context of the reduced parameterization of a VPA or a separable VPA, the catch-at-age data are not sufficient. There has been a growing recognition of the limitation of the assumptions of the traditional VPA approaches, particularly in situations where large uncertainties exist in the catch-at-age data. This has led to increasing model complexity and greater emphasis on a proper statistical framework. Recently, better models have been developed (most notably the MULTIFAN-CL approach, Fournier et al 1998). These general approaches are referred to here as SCALIA (statistical catch-at-age/length integrated analysis) models.

SCALIA models allow sampling variability in the catch-at-age data. Instead of full separability between age- and year effects in the fishing mortality, smoothness assumptions over ages and over years are imposed on these parameters. The SCALIA approach also attempts to ground the assessment models in a more formal statistical framework by developing likelihood functions. Since no additional data are being introduced, the imposed smoothness assumptions are critical. These assumptions are introduced as "penalty" terms in the likelihood function, for example by a third difference penalty on $F_{i,t}$ in age for given year and a corresponding penalty term to constrain temporal variability in $F_{i,t}$ for given age. In addition to the process of mortality, the process of reproduction (recruitment) needs to be modeled to achieve a tractable model.

Taken together, the SCALIA approach results in a penalized likelihood function which may include at least 6 basic components: (1) a likelihood for the predicted total catch; (2) a likelihood for the predicted proportion of the catch by age/size; (3) a likelihood for the predicted relative abundance indices; (4) a penalty for the smoothness in $F_{i,t}$ between ages within a year; (5) a penalty to constraint temporal variability in $F_{i,t}$ and (6) a penalty to constrain the amount of variability in the recruitment relative to the specified stock-recruitment relationship. Simply finding the maximum solution to such a complex penalized likelihood function constitutes a substantial, but solvable, computational problem, often using the technique of automatic differentiation (Fournier et al 1998). However, understanding the statistical properties of the estimates derived from such models represents a much larger and critical challenge if the model results are to underlie management decisions. This might require innovative analytical approaches as well as simulation testing given the complexity of the models.

3. SBT Application

Southern bluefin tuna (SBT) is a long-lived (up to 40^+ years), highly mobile, widely distributed fish found throughout most of the southern temperate ocean. It is among the most highly valued tuna for the Japanese sashimi market and has been extensively exploited since 1960. The stock is considered severely depleted with the current parental biomass estimated at 6-15% of its pre-exploitation size (Anon., 1998). Restrictive and large catch reductions were imposed in 1989, but the parental stock is estimated to have continued to decline and catches in recent years have increased due to unregulated components of the fishery. Large uncertainty and disagreement exists regarding whether the stock will recover or continue to decline under current catch levels.

Since 1982, the SBT assessments have been based on traditional VPA. However, the SBT assessments are characterized by large uncertainties in the input data and biological parameters (Polacheck et al, 1999). The catch-at-age estimates are derived from converting estimates of the length frequency of the catch into age distributions based on estimates of estimated growth curves from tag recapture data. In addition, inconsistencies exist among the relative abundance indices for the different age-classes if they are assumed linearly related to abundance. An initial application of a SCALIA model has been developed (Butterworth et al, 2000) which attempts to provide a consistent interpretation of all the available data. However, the application of the general SCALIA approach both to SBT and in general presents a number of unresolved challenging statistical problems.

4. Statistical Challenges

1. *The development of appropriate likelihood functions*: The primary "data" being fitted in the SBT stock assessment are estimates of the aggregated catch-at-age data and age-specific relative abundance indices. Simplified likelihood functions are used for these two sets of "data". The relative indices are treated as independent and log-normally distributed. The catch-at-age proportions are treated as multinomial. The age-specific relative indices are derived from GLM modeling of catch and effort aggregated into monthly geographic strata in which the age distribution is estimated from (small) samples of the actual catch or by extrapolation from near by strata. The distributional assumptions are clearly inappropriate, and statistical challenges in using these data are numerous: What is a more

appropriate formulation for likelihood functions? What are the consequences of ignoring the lack of independence among the different age-specific indices? The process, which generates the catch-at-age data, has multiple stages and is not a simple multinomial one. Can the actual collection and compilation of the catch-at-age data (at least in the future) be put into a rigorous statistical framework to allow development of a realistic likelihood function? If a multinomial distribution is used to approximate the actual likelihood, how should the "effective" sample size be determined? How should differences in sampling over time be adequately modeled in terms of changes in effective sample sizes? Large changes in key output statistics of interest for SBT-management have, in fact, been experienced with changes in the assumed effective sample sizes from 40 to 100.

2. *Weightings of penalty terms*: Each of the penalty terms imposed on the model requires a relative weighting of substantial impact on the parameter estimates and model predictions. Their specification is usually relatively arbitrary depending upon the user's *a priori* judgement about the degree of variability in various aspects of the population and fishery dynamic processes compared to the sampling error in the observed data. These penalty weights should ideally be based on an understanding of the various sources of variation in the data. This understanding needs to be improved.

3. *Over-parameterization and measures of lack of fit*: Even with smoothness constraints, the SCALIA models are often over-parameterized. There might be from hundreds to several thousands free parameters. From a statistical perspective, are such large numbers of parameters justifiable? Can equally parsimonious models be developed with substantially fewer parameters? Given the multi-component penalized likelihood function, what statistical (probabilistic) criteria are appropriate for judging whether a given model is over-parameterized relative to simpler alternatives? What are appropriate diagnostics for assessing potential lack of fit in such complex models? A problem that has been encountered in the SBT assessment is that strong temporal trends can be encountered in the residual patterns for one set of inputs. How should results from such models be considered? When should lack of fit in one component of a multi-component fitting procedure be considered to represent an inappropriate combination of model choice and parameterizations? This type of lack of fit might be overcome by introducing additional assumptions and parameters. In the SBT situation, Butterworth et al (2000) introduced an auto-correlation structure for this purpose, but it altered the results considerably. Is the use of such an auto-correlation structure statistically appropriate or does it conceal a more fundamental model mis-specification?

4. *Integration across alternative constraints and parameterizations*: Model results (particularly those critical for providing management advice) can be sensitive to alternative formulations including different penalty functions and basic parameterizations. For managers, some form of integration across the various alternatives is important so that overall assessment of uncertainties and risk associated with possible management actions can be evaluated. A Bayesian approach might be viable for integrating across various alternatives and uncertainties. However, the specification of priors (especially joint priors) is not straight forward, nor is it simple to calculate the posterior distribution for such high dimensional parameter models. Moreover, in other cases, different parameterizations are not nested and the penalized likelihood functions may contain differing numbers of components. In these cases, integrating across the different model formulations represents a major, largely unaddressed statistical challenge.

There is thus a large scope for improving the statistical foundation of fishery stock assessments. However, it would be naïve to think that the solution to the basic estimation problem is simply a statistical one, no matter how interesting the statistical challenges. The limitations in the currently existing data, in our ability to measure the important quantities of interests and in the basic understanding of the fish stock and fishery dynamics, must be acknowledged and recognized. Working to solve these limitations is a challenge requiring the collaborative, creative interaction of statisticians, fishery biologists and assessment modelers. This will entail developing a mutual understanding of the detailed issues combined with a willingness to go beyond traditional disciplinary boundaries.

REFERENCES

Anon. 1998. Report of the 1998 Scientific Committee meeting. 3-6 August (1998). Tokyo, Japan. Commission for the Conservation of Southern Bluefin Tuna (CCSBT).

Butterworth D. S., J. N. Ianelli, and R. Hilborn. (2000). A statistical time-series model for stock assessment of southern bluefin tuna. CCSBT-SAP/0005/9.

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

Polacheck T. W., A. Preece, N.Klaer. and A. Betlehem (1998). Treatment of data and model uncertainties in the assessment of southern bluefin tuna stocks. In: F. Funk, T.J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang (eds.), Fishery stock assessment models. University of Alaska Sea Grant, ASG-98-01.

United Nations. (1995). Agreement for the implementation of the provisions of the United Nations Convention on the Law of the Sea of 10 December 1982 relating to the conservation and management of straddling fish stocks and highly migratory fish stocks

ACKNOWLEDGEMENTS

The ideas in this paper have been developed from wide ranging discussions with numerous statisticians and stock assessment modelers. In particular, I would like to acknowledge Dale Kolody and Ann Preece.

RÉSUMÉ

Les «meilleures» estimations sur l'état d'un stock et sur les conséquences d'actions ultérieures ne constituent plus des bases suffisantes pour la constitution de conseils de gestion scientifique en matière de pêche. Selon les normes internationales actuelles, il est du ressort de la direction de s'assurer du faible risque de surexploitation halieutique. Ceci exige un calcul précis de l'incertitude relative aux quantités appropriées estimées. Les évaluations des principales ressources halieutiques sont basées sur des estimations historiques de la taille des différentes populations de poissons, d'après des analyses de données de l'âge au moment de la pêche, ajoutées à des indices d'abondance relative et/ou des variations d'efforts halieutiques.

Ce problème d'estimation est surdéterminé. Afin de le résoudre, il est nécessaire d'ajouter aux paramètres des suppositions et/ou des contraintes. Les méthodes analytiques utilisées dans le secteur de la pêche en vue de définir le problème dans un cadre statistique approprié, tout en apportant un réalisme et une souplesse accrus aux contraintes et aux suppositions paramétriques, sont devenues de plus en plus complexes ces dernières années. Les modèles les plus récents adoptent une structure de probabilité de pénalités (souvent dans un contexte bayésien) intégrant un certain nombre de sources de données disparates et portant sur l'estimation de centaines (et dans certains cas de milliers) de paramètres (souvent fortement liés). L'application de ces modèles aux stocks de poissons existants comporte de nombreux défis statistiques, en particulier en ce qui concerne l'analyse quantitative de l'incertitude sous-jacente. Parmi ces difficultés on compte: (1) le développement de fonctions de probabilités appropriées, (2) les coéfficients à appliquer aux termes de pénalités, (3) la paramétrisation excessive et des mesures de manque de correspondance et (4) intégration couvrant les contraintes et les paramétrisations possibles. Les questions et les problèmes en cause sont développés dans le contexte des méthodes évolutives utilisées pour l'évaluation du statut du thon à aileron bleu.