



Summary of revisions to the tagging likelihood component of the CCSBT operating model

Paige Eveson

**Prepared for the CCSBT 14th Meeting of the Extended Scientific Committee 5-11
September 2009
Busan, Korea**

Table of Contents

Introduction.....	2
Methods.....	2
Tagging model	2
Predicted Tag Returns	4
Tag likelihood	4
Standardized residuals	5
Mixing assumption.....	6
Summary	7
References.....	7

Introduction

In previous versions of the CCSBT operating model (OM) (i.e., up to and including version sbtmod21), the 1990s RMP tag release and recapture data had been modelled by pooling the data over cohorts, and not keeping track of recaptures by the year of release. The SBT tagging experiments were designed so that cohorts were tagged in multiple consecutive years, and as such the recapture data contain information on both fishing mortality and natural mortality. By pooling the data across cohorts, the information on natural mortality is greatly reduced.

The most recent version of the OM (sbtmod22) includes a revised model for the tag data based on a Brownie model (Brownie et al. 1985). Brownie models were designed for analysing multiyear tag data, and can provide direct estimates of both natural and fishing mortality rates. The ability of the Brownie approach to separate natural and fishing mortality rates is a direct result of the multiple release events.

This document describes the revised model and likelihood for the 1990s tagging data.

Methods

Tagging model

The dynamics for fish that are tagged and released are assumed to be the same as those for the general population. Thus, the tagging model assumes two seasons per year, one from Jan 1 to Jun 30 and the other from Jul 1 to Dec 31, with pulse fisheries operating at the start of each season. Tag releases have generally occurred near the beginning of the calendar year, so they are treated as discrete annual events occurring at the start of the first fishing season (i.e., on Jan 1). Because newly tagged fish will not be completely mixed during the season following their release, the model allows for the harvest rate to differ between tagged fish in the season directly following their release and untagged fish in that same season.

We know that a significant proportion of tags recaptured in the fisheries are not returned. Thus, age and year-specific reporting rate estimates are included in the model. A number of alternative estimates for the reporting rates have been proposed in past, but the estimates currently being used in the OM are from Table 5a, Option 8, of Eveson and Polacheck (2005), adjusted for the various overcatch scenarios.

Tag shedding is another issue that needs to be taken into account. All fish were double tagged so that shedding rates could be estimated (based on the number of recaptured fish with one tag versus two tags still attached). We assume that the probability of a tag being retained after time τ (in years) at liberty can be described by

$$Q(\tau) = \xi e^{-\Omega\tau}$$

where ξ is the fraction of tags immediately retained (i.e., proportion $1 - \xi$ are immediately shed) and Ω is the continuous shedding rate.

In an independent analyses preformed by Dr. W.S. Hearn (CSIRO Marine and Atmospheric Research), shedding rates were found to be tagger-dependent, so separate values of ξ and Ω were estimated for 6 groups of taggers found to have statistically similar shedding rates (Eveson and Polacheck 2009).

Taking into account all of the above, the probability of a fish from cohort c tagged at age a by a tagger in group g being recaptured at age i and having *at least one* of its two tags returned is:

$$P_{c,a,g,i} = \begin{cases} 0 & i < a \\ (2\xi_g f'_{c,g,i} - \xi_g f''_{c,g,i}) v_{c,i} & i = a \\ (2\xi_g S'_{c,g,a} f'_{c,g,i} - \xi_g^2 S''_{c,g,a} f''_{c,g,i}) v_{c,i} & i = a+1 \\ \left(2\xi_g S'_{c,g,a} \left(\prod_{j=a+1}^{i-1} S'_{c,g,j} \right) f'_{c,g,i} - \xi_g^2 S''_{c,g,a} \left(\prod_{j=a+1}^{i-1} S''_{c,g,j} \right) f''_{c,g,i} \right) v_{c,i} & i > a+1 \end{cases}$$

where

$$\begin{aligned} S'_{c,g,i} &= (1 - h_{1,c,i})(1 - h_{2,c,i}) \exp(-M_i - \Omega_g) \\ S''_{c,g,i} &= (1 - h_{1,c,i})(1 - h_{2,c,i}) \exp(-M_i - 2\Omega_g) \\ f'_{c,g,i} &= h_{1,c,i} + (1 - h_{1,c,i}) \exp(-0.5(M_i + \Omega_g)) h_{2,c,i} \\ f''_{c,g,i} &= h_{1,c,i} + (1 - h_{1,c,i}) \exp(-0.5(M_i + 2\Omega_g)) h_{2,c,i} \\ S'^*_{c,g,i} &= (1 - h^*_{1,c,i})(1 - h_{2,c,i}) \exp(-M_i - \Omega_g) \\ S''^*_{c,g,i} &= (1 - h^*_{1,c,i})(1 - h_{2,c,i}) \exp(-M_i - 2\Omega_g) \\ f'^*_{c,g,i} &= h^*_{1,c,i} + (1 - h^*_{1,c,i}) \exp(-0.5(M_i + \Omega_g)) h_{2,c,i} \\ f''^*_{c,g,i} &= h^*_{1,c,i} + (1 - h^*_{1,c,i}) \exp(-0.5(M_i + 2\Omega_g)) h_{2,c,i} \end{aligned}$$

Parameters are defined as follows:

- M_i is the natural mortality rate for age i fish
- $h_{1,c,i}$ is the proportion of age i fish from cohort c harvested in season 1
- $h_{2,c,i}$ is the proportion of age i fish from cohort c harvested in season 2
- $h^*_{1,c,i}$ is the proportion of age i fish from cohort c that were tagged at age i and recaptured in the season directly following release
- $v_{c,i}$ is the reporting rate for age i fish from cohort c
- ξ_g is the immediate retention rate for tags released by tagger group g
- Ω_g is the continuous shedding rate for tags released by tagger group g

In terms of the parameterization used for the population model of the OM:

$$h_{1,c,i} = \sum_{f \in f^1} H_{f,c+i,i}$$

$$h_{2,c,i} = \sum_{f \in f^2} H_{f,c+i,i}$$

where $H_{f,c+i,i}$ is the fishing proportion of fishery f in year $c+i$ for fish of age i (i.e., for fish of age i from cohort c).

Predicted Tag Returns

Let $N_{c,a,g}$ denote the number of fish from cohort c tagged at age a by taggers in group g . We will refer to this set of tag releases as set (c, a, g) . Let $R_{c,a,g,i}$ be the observed number of fish from release set (c, a, g) that were recaptured at age i and had at least one of their tags returned (for simplicity, we will refer to this as the number of tag returns). Then, the predicted number of tag returns is given by

$$\hat{R}_{c,a,g,i} = N_{c,a,g} p_{c,a,g,i}$$

Tag likelihood

If all assumptions of a Brownie tagging model are met (e.g., complete mixing; independence between tagged fish), then the numbers of tags returned at ages a to I , plus the number not returned by age I , corresponding to the $N_{c,a,g}$ releases from release set (c, a, g) have a multinomial distribution; i.e.,

$$\{R_{c,a,g,a}, \dots, R_{c,a,g,I}, N_{c,a,g} - R_{c,a,g,\bullet}\} \sim \text{Multinom}\left(N_{c,a,g}, \{p_{c,a,g,a}, \dots, p_{c,a,g,I}, 1 - p_{c,a,g,\bullet}\}\right)$$

where a dot in the subscript denotes summation over the index it replaces (e.g.,

$$R_{c,a,g,\bullet} = \sum_{i=a}^I R_{c,a,g,i}.$$

However, in practice, the tag return data will almost certainly be over-dispersed relative to a multinomial distribution (i.e., more variable). To account for this, we model the tag returns for release set (c, a, g) using a Dirichlet-multinomial distribution, parameterized such that the amount of variance in the data is φ times that of multinomial data (for details refer to Polacheck et al. 2006). Then, the likelihood function for the observed numbers of returns from all release sets is the product of Dirichlet-multinomials.

Specifically,

$$L_R = \prod_c \prod_g \left\{ K_{c,g} \prod_a \left(\frac{\Gamma(\omega_{c,a,g})}{\Gamma(N_{c,a,g} + \omega_{c,a,g})} \frac{\Gamma(R'_{c,a,g,\bullet} + \omega_{c,a,g} p'_{c,a,g,\bullet})}{\Gamma(\omega_{c,a,g} p'_{c,a,g,\bullet})} \prod_{i=a}^I \frac{\Gamma(R_{c,a,g,i} + \omega_{c,a,g} p_{c,a,g,i})}{\Gamma(\omega_{c,a,g} p_{c,a,g,i})} \right) \right\}$$

where:

$$\omega_{c,a,g} = (N_{c,a,g} - \varphi) / (\varphi - 1)$$

$$R'_{c,a,g,\bullet} = N_{c,a,g} - R_{c,a,g,\bullet}$$

$$p'_{c,a,g,\bullet} = 1 - p_{c,a,g,\bullet}$$

$$K_{c,g} = \prod_a \frac{N_{c,a,g}!}{\prod_{i=a}^I R_{c,a,g,i}! (N_{c,a,g} - R_{c,a,g,\bullet})!}$$

Note that $K_{c,g}$ is a constant that can be omitted when maximizing the likelihood.

The negative log likelihood (leaving off the constant) can then be expressed as

$$\begin{aligned} \ln L_R = & \sum_c \sum_g \sum_a \sum_{i=a}^I \left(\ln \Gamma(R_{c,a,g,i} + \omega_{c,a,g} p_{c,a,g,i}) - \ln \Gamma(\omega_{c,a,g} p_{c,a,g,i}) \right) + \\ & \sum_c \sum_g \sum_a \left(\ln \Gamma(\omega_{c,a,g}) - \ln \Gamma(N_{c,a,g} + \omega_{c,a,g}) + \ln \Gamma(R'_{c,a,g,\bullet} + \omega_{c,a,g} p'_{c,a,g,\bullet}) - \ln \Gamma(\omega_{c,a,g} p'_{c,a,g,\bullet}) \right) \end{aligned}$$

Finally, the negative log likelihood for the tagging model is

$$-\ln L_{tag} = -\ln L_R$$

In optimizing the likelihood, we input the reporting rate estimates and shedding parameter estimates into the model as known. Also, we have found in simulations that the over-dispersion factor ϕ cannot be estimated reliably within the model. Thus, we estimate this parameter based on the residuals from the model assuming multinomial returns, as described in the ‘Standardized residuals’ section below, and input it as fixed to the model.

Standardized residuals

Under the assumption of multinomial tag return data, the observed number of tag returns at age i corresponding to release set (c, a, g) is approximately normally distributed with mean and variance as follows:

$$R_{c,a,g,i} \sim \text{Normal}\left(N_{c,a,g} p_{c,a,g,i}, N_{c,a,g} p_{c,a,g,i} (1 - p_{c,a,g,i})\right)$$

Thus, approximate standardized normal residuals can be calculated as

$$\frac{R_{c,a,g,i} - N_{c,a,g} p_{c,a,g,i}}{\sqrt{N_{c,a,g} p_{c,a,g,i} (1 - p_{c,a,g,i})}}$$

If the multinomial assumption is correct, the variance of these standardized residuals should be approximately 1. If the variance is in fact x , this provides a reasonable estimate of the over-dispersion factor (i.e., $\hat{\phi} = x$).

Thus, standardized residuals for the Dirichlet-multinomial model can be calculated as

$$\frac{R_{c,a,g,i} - N_{c,a,g} P_{c,a,g,i}}{\sqrt{\hat{\phi} N_{c,a,g} P_{c,a,g,i} (1 - P_{c,a,g,i})}}$$

An estimate of ϕ was calculated using the tag residuals obtained from each cell of the grid under the assumption of multinomial tag returns, and was found to range from 2.2 to 2.9 with a mean of 2.35. (Note that this analysis was performed using the 3 January 2009 version of sbtmod22, and could be updated using the most recent version.)

Mixing assumption

Incomplete mixing of tagged animals within the full population and heterogeneity in the capture probabilities is a common and challenging problem with tagging experiments. We classify incomplete mixing into two basic types: systematic and unsystematic.

Unsystematic incomplete mixing refers to situations where there is still a large degree of mixing among tagged and untagged fish and the pattern of mixing has a large random component such that, on average, the probability of capture is the same for tagged and untagged fish. This type of non-mixing results in overdispersion in the tag return data, and we dealt with it by using a Dirichlet-multinomial distribution for the tag returns instead of a multinomial distribution.

Systematic incomplete mixing refers to the situation where there is a systematic and repeatable pattern of non-mixing between tagged and untagged fish – for example, if all tagging was done late in the season in one location and fish in that location and time period only remain in one part of the stock's overall range. Such systematic non-mixing can induce biases into the population and mortality estimates if not accounted for within the model. One type of systematic non-mixing can occur if tagging occurs around the same time as peak catches, in which case it is important to allow for an initial period of non-mixing. The revised tag model does so by allowing for the harvest rate to differ between tagged fish in the season directly following their release and untagged fish in that same season. Another type of systematic non-mixing may occur for SBT due to the fact that tagging occurs almost exclusively off the south coast of Western Australia and in the Great Australian Bight. As such, it is possible that only a fraction of the overall juvenile population is available to tagging. If the component of the population not available to tagging remains in a different geographical region than the tagged fish, then this could result in biased estimates of the mortality rates (Polacheck et al. 2002 use simulations to explore the nature and degree of biases that can result from such non-mixing). It may be possible to allow for such non-mixing in the OM in a crude but simple way by assuming that fishing mortality rates on the tagged population differ by a constant factor from fishing mortality rates on the general population. This is currently being explored by members of the CCSBT MP Technical Group.

Summary

In this paper we described the revised tagging model and likelihood being used in the most recent version of the CCSBT OM (sbtmod22, 21 July 2009). The SBT tagging experiments were designed so that cohorts were tagged in multiple consecutive years (i.e., at consecutive ages). The revised tag model is based on a Brownie model that takes advantage of the information on natural mortality provided by this experimental design. In the previous tag model, the tag data were pooled across cohorts so that the information on natural mortality was greatly reduced. In addition, the revised model incorporates tagger-specific estimates of shedding rates that were estimated from an independent analysis of the double-tagging data. Tag shedding was not taken into account in the previous tag model.

References

- Brownie, C., Anderson, D.R., Burnham, K.P., and Robson, D.S. 1985. Statistical inference from band recovery data: a handbook. US Fish Wildl. Resour. Publ. 156.
- Eveson, P. and Polacheck, T. 2009. Analyses of tag return data from the CCSBT SRP tagging program – 2009. CCSBT-ESC/00909/XX.
- Polacheck, T., Eveson, J.P., and Laslett, G.M. 2002. Tagging in a spatial context: design and analysis considerations. Appendix 12 of Estimation of mortality rates from tagging data for pelagic fisheries: analysis and experimental design. Final Report to the Fisheries Research and Development Corporation. FRDC Project 2002/015. ISBN 1 921061 03 0.
- Eveson, P. and T. Polacheck. 2005. Updated estimates of tag reporting rates for the 1990s tagging experiments. CCSBT-MPTM/0502/05.
- Polacheck, T., Eveson, J.P., Laslett, G.M., Pollock, K.H., and Hearn, W.S. 2006. Integrating catch-at-age and multiyear tagging data: a combined Brownie and Petersen estimation approach in a fishery context. Can. J. Fish. Aquat. Sci. 63: 534-548.