CCAMLR Working Group Document Submission Form

To be completed by the Secretariat:	
Document No.:	**WG-SAM-16/**
Date submitted:	
Original Language:	English
To be completed by the author:	
Meeting:	WG-SAM-2016
Agenda Item No(s):	2.1

Title **Performance metrics to index the spatial coverage of mark recapture data**

Author(s) C. Marsh, A. Dunn, S. Mormede.
Address(s) National Institute of Water and Atmospheric Research (NIWA) Ltd
Private Bag 14901, Wellington, New Zealand

Name and email address of person submitting paper: *rohan.currey@mpi.govt.nz*

Published or accepted for publication elsewhere? Yes [ ] No [X]

If published or in press, give details:

To be considered for publication in *CCAMLR Science*?[1] Yes [ ] No [X]

[1] By indicating that the paper is to be considered for publication in *CCAMLR Science*, the authors have agreed that the paper can be considered by the Editorial Board of the journal and that, if the paper is accepted for peer review, it is the responsibility of the authors to ensure that permission to publish data and cite unpublished working group papers has been received.

ABSTRACT

Understanding the spatial distribution of the release of tagged fish and the subsequent spatial coverage of fishing effort to recapture tagged fish is an important consideration when interpreting the biomass estimated from mark recapture data. In 2015, the Scientific Committee identified that measures of spatial overlap and potential bias in the development of tag-based biomass estimates are an important focus topic for WG-SAM (SC-CAMLR-XXXIV, 2015, para 3.83). This paper outlines developments towards new biomass-weighted spatial overlap summary statistics for tagging data. The method calculates two statistics: (i) a single measure of the degree of spatial overlap between the release of tagged fish and subsequent fishing effort for tag recovery, labelled the tag spatial overlap (TSO) statistic and (ii) a measure of the potential bias in the biomass estimate calculated from non-homogenous spatial mark-recapture data, labelled the tag spatial bias (TSB) statistic.

We apply the method to a single case study area, SSRUs 88.2H to illustrate its use. We found that the median tag spatial overlap statistic in 2012 was 70% (95% CIs 53–80%), and the median tag overlap bias was 88% (95% CIs 64–114%). In 2014 the median tag spatial overlap statistic 60% (95% CIs 50-69%), and the median tag overlap bias was 192% (95% CIs 141-268%).

The TSO statistic and TSB statistic provide a simple; and intuitive approach to indexing the degree of spatial overlap and potential bias from mark recapture data, with few assumptions required on the distribution of fish, movement, or catch history. Potential uses of these statistics could include a measure of the total spatial distribution of effort through time in research or developing fisheries, and a measure of the change in bias of mark-recapture estimates through time.

Marsh, C., Dunn, A., Mormede, S. (2016). Performance metrics to index the spatial coverage of mark recapture data. Document WG-SAM-16/XX. CCAMLR, Hobart, Australia. 10 p.

1. INTRODUCTION

Understanding the spatial distribution of the releases of tagged fish and the subsequent spatial coverage of fishing effort to recapture tagged fish is an important consideration when interpreting the biomass estimated from mark recapture data. The Scientific Committee identified that measures of spatial overlap and potential bias in the development of tag-based biomass estimates were an important focus topic for WG-SAM (SC-CAMLR-XXXIV, 2015, para 3.83). They requested advice on methods to quantify the level of spatial overlap between tagged fish and subsequent fishing effort, and evaluation of the potential bias introduced into stock assessments and tag-based biomass estimates when the distributions of tagged fish, fishing effort and the underlying stock distribution are spatially heterogeneous.

This paper outlines developments towards a new method that describes biomass-weighted spatial overlap summary statistics for areas using mark-recapture data. The method calculates two statistics: (i) a single measure of the degree of spatial overlap between the release of tagged fish and subsequent fishing effort, labelled the TSO statistic and (ii) a measure of the potential bias in the biomass estimate calculated from non-homogenous spatial mark recapture data, labelled the TSB statistic.

The method uses an analysis that integrates data on the release of tagged fish, the recapture of tagged fish and relative abundance from, for example, catch rate data from a fishery or a survey. It evaluates the quality of the spatial overlap of these data and estimates a measure of potential spatial bias in these data. The method is implemented using a simple latent variable model. The method uses an integrated analysis that can be easily modified to combine data from a variety of sources to inform the estimation of parameters (Fournier & Archibald, 1982; Methot, 1990).

The TSO statistic can be considered a measure of the potential vulnerable biomass accounted for by tagging data. The TSO is expressed as a percentage of the total vulnerable biomass within the area fished (it does not take into account potential vulnerable biomass outside of the area fished). The TSB statistic is closely related, and can be considered a measure of the potential bias in the Chapman mark-recapture estimator for the area fished (Chapman, 1951). We use the mark-recapture data from SSRUs 88.2H from two contrasting years to illustrate the method, and compare how the statistics perform with a moderate amount of tagging data in a high overlap scenario (SSRU 88.2H in 2012) and a case with a lower overlap (SSRU 88.2H in 2014).

2. METHODS

Introduction

The method considers a region as a series of independent cells, where each cell is small enough to be considered an area where there is homogenous mixing of tagged and untagged fish, and no mixing of fish between cells over the time span studied (here, we assume that the cell size was such that there was no mixing of tagged fish between cells after only one year at liberty). Within the region, we assume that that each homogenous cell i has some unknown vulnerable biomass θ_i. Then, for each cell in the model we require a measure of relative vulnerable abundance $O_{c,i}$, and, in those cells with at least one recapture, a local cell-based Chapman estimate of vulnerable biomass $O_{t,i}$.

We assume that the measure of relative vulnerable abundance $O_{c,i}$ (typically standardised catch rates) is proportional to abundance across the cells for some subset of the available data. However, as the relationship between catch rates and cell abundance may be different across factors where data are insufficient for standardisation (for example vessel, gear, or even season), we could instead define a set of observations $O_{c(x),i}$ for each factor. For example, we could define a relative catch rate for each gear type g vessel $O_{c(g),i}$, each vessel v as $O_{c(v),i}$ or even vessel/gear combinations $O_{c(v,g),i}$ for the model.

To supplement the catch rate data, we require a measure of available fishable area to transform the relative density of the catch rate data for each cell into a measure of relative abundance. Hence for each cell we assume either that the entire cell is of a depth that is vulnerable to fishing, or otherwise we estimate the seabed area of each cell vulnerable to fishing. In this instance we assume that the cells are small enough to mostly comprise only fishing depths, and therefore seabed area is ignored.

For some subset of cells, we require some mark recapture observations. These can be used to calculate a local cell-based Chapman estimate of biomass $O_{t,i}$. We assume the tagged fish released in a cell are available to be recaptured in the following year in that same cell, after discounting for initial tagging mortality, ongoing tag loss and natural mortality.

An example of this schematic is shown below in Figure 1.

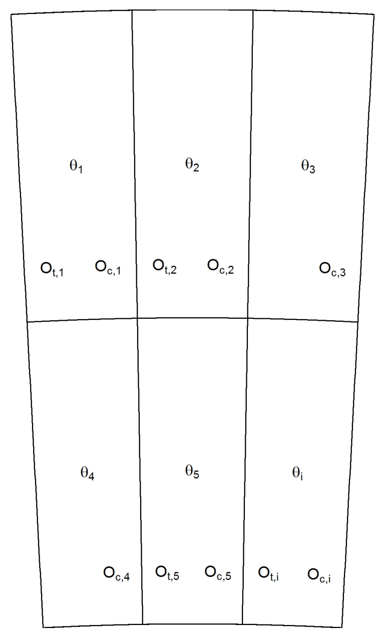

Figure 1: Defining the state, where θ_i is the unknown biomass in cell i, $O_{c,i}$ are the CPUE observations (defined in this example figure as available cells 1, 2, 3, 4, and 5) and $O_{t,i}$ are the mark recapture biomass observations (defined in the example as available for cells 1, 2, and 5).

The tag spatial overlap (TSO) statistic

If the vulnerable biomass in cell i (θ_i) can be estimated ($\hat{\theta}_i$), then we can define the TSO statistic as the ratio of the estimated vulnerable biomass formed from the tagging data summed over cells that have at least one recapture, divided by the estimated vulnerable biomass summed over all cells (including those cells where there were no tag recaptures):

$$\mathrm{TSO} = \frac{\sum_{i \in O_t} \hat{\theta}_i}{\sum_i \hat{\theta}_i} \times \frac{100}{1} \tag{2.1}$$

where $\sum_{i \in O_t} \hat{\theta}_i$ is the sum of the estimated biomass from the cells where we have had at least one mark recapture, and $\sum_i \hat{\theta}_i$ is the sum of the estimated biomass from all cells using another abundance estimate (here from catch rate data).

We can derive an observation of the biomass in the individual cells that have had a tagged fish recaptured using the Chapman estimator (Burch, Parker, & Welsford, 2015; Hillary, 2009):

$$O_{t,i} = \frac{\left(\tilde{n}_{(y-1),i} + 1\right)\left(c_{y,i} + \delta\right)}{\left(m_{y,i} + 1\right)} - \delta \tag{2.2}$$

Where

$\tilde{n}_{(y-1),i}$ = total number of fish tagged and released in the previous year (i.e., *y-1*) that were estimated to be available for recapture in year *y* and cell *i,*
$c_{y,i}$ = total weight of all fish caught (i.e., examined for presence of tags) in year *y* and cell *i,*
$m_{y,i}$ = total number of tagged fish released in year (*y-1*) and recaptured in year *y* and cell *i*, and
δ = the mean weight of an individual fish.

Note that the number of fish tagged and released in year (*y-1*) available for recapture can be estimated using the tag loss rate and the rate of natural mortality:

$$\tilde{n}_{(y-1),i} = n_{(y-1),i}\left(1 - t\right)e^{-(M+f)} \tag{2.3}$$

Where

t = the initial tag loss proportion,
f = ongoing tag loss rate, and
M = natural mortality.

The TSO statistic describes how much of the estimated biomass is informed by the mark recapture data, divided by the total biomass in all cells estimated proportionally to CPUE, and hence will always be a value between zero and 100.

The tag spatial bias (TSB) statistic

Similar to the TSO, the TSB statistic can be calculated as the ratio of the biomass estimate from the mark recapture data from all cells treated as a single cell ($\hat{B}_t$), divided by the total biomass $\sum_i \hat{\theta}_i$ calculated as above, i.e.,

$$\text{TSB} = \frac{\hat{B}_t}{\sum_i \hat{\theta}_i} \times \frac{100}{1} \tag{2.4}$$

And

$$\hat{B}_t = \frac{\left(\tilde{n}_{(y-1)} + 1\right)\left(c_{\mathrm{y}} + \delta\right)}{\left(m_y + 1\right)} - \delta \tag{2.5}$$

Where

$\tilde{n}_{(y-1)}$ = total number of fish tagged and released in the previous year (i.e., *y*-1) that are estimated to be available for recapture in year *y,*

$c_{y,i}$ = total weight of all fish caught (i.e., examined for presence of tags) in year *y*,,

$m_{y,i}$ = total number of tagged fish released in year (*y-1*) and recaptured in year *y*, and

δ = the mean weight of an individual fish.

The TSB statistic describes how different the Chapman estimate of biomass for the entire area is from the spatially explicit estimate of biomass. It is a positive value, less than 100% if the Chapman estimator under-estimates the underlying biomass, and higher than 100% otherwise.

Estimation

We use a latent variable model with an integrated analysis approach to determine the TSO and TSB statistics. We estimate θ_i in each cell by using the observations from the mark recapture data $O_{t,i}$ and $O_{c,i}$ to inform the estimation. We use equation (2.3) in cells where there are mark-recapture data available;

$$\begin{aligned} E\left[\theta_i\right] &= O_{t,i} \\ &= \frac{\left(n_{(y-1),i} + 1\right)\left(c_{\mathrm{y},i} + \delta\right)}{\left(m_{y,i} + 1\right)} - \delta \end{aligned} \tag{2.6}$$

In cells where there are no mark recapture data but catch rate information we use the following equation:

$$E\left[\theta_i\right] = qO_{\mathrm{c},i}A_i \tag{2.7}$$

Where q is a scaler (catchability coefficient) and A_i is the area of cell i to convert the catch rate from density to abundance.

We assume that both the expectations in equations (2.6) and (2.7) are lognormally distributed. For catch rate observations, the sampling error (CV) derived by bootstrapping catch rate data was calculated, except in cells that contained only a single set or fishing event. For single event situations we imputed the CV as the mean CV over all other cells. For the Chapmans biomass estimated observations, we assumed that the sampling error was equal to the analytical estimate of the variance using the usual Chapman variance equation (Burch et al., 2015). An alternative to the analytical CV would be to derive a bootstrapped estimate.

Estimable parameters were, the biomass in each cell (θi) and the relative catchability q for the catch rate observations. Priors for vulnerable biomass θ across all cells were assumed to be uniform with no bounds, and the prior for the catchability coefficient assumed to be log-uniform. Parameter estimates were made using a naive Metropolis Independence sampler (Hastings, 1970) with an adaptive proposal distribution. The negative log likelihood is as follows.

$$
\begin{aligned}
-\ln(L) = &\sum_{i \in O_t} \left(\log\left(\sigma_{t,i}\right) + 0.5\left(\frac{\log\left(\hat{\theta}_i / O_{t,i}\right)}{\sigma_{t,i}} + 0.5\sigma_{t,i} \right)^2 \right) \\
&+ \sum_{\forall i} \left(\log\left(\sigma_{c,i}\right) + 0.5\left(\frac{\log\left(\hat{\theta}_i / O_{c,i}\right)}{\sigma_{c,i}} + 0.5\sigma_{c,i} \right)^2 \right)
\end{aligned}
\tag{2.8}
$$

Where,

$$\sigma_{t,i} = \sqrt{\ln(CV_{t,i}^2 + 1)}$$

$$\sigma_{c,i} = \sqrt{\ln(CV_{c,i}^2 + 1)}$$

The objective function which is evaluated in the MCMC routine also includes prior contributions and penalties for zero and non-positive estimate values. We ran MCMCs for 600 000 iterations with a burn in of 60 000 and thinning every 100 iterations.

3. CASE STUDY

We use the mark-recapture and catch rate data from SSRU 88.2H from two fishing years, 2012 and 2014, to illustrate the method. Figure 2 shows the seamounts in SSRU 88.2H that were chosen as the spatial cells. These were previously defined by Parker (2014). The spatial mark-recapture data are shown in Figure 3.

Values for parameters driving the tagged population available after one year at liberty (see Equation (2.3)) are defined as, $t = 0.1$, $M = 0.13$ y^{-1}, and f = 0.066 y^{-1} following Mormede et al. (2014).

The catch limit in SSRU 88.2H was reduced from 406t in 2012 to 200t in 2014, which resulted in a 67% drop in effort (Figure 3). This reduction in effort translated into a decrease in the number of recaptured fish observed between 2012 and 2014 (right panels), and a decrease in the spatial coverage of the effort, suggesting the tag overlap statistic for 2014 should be lower than for 2012.

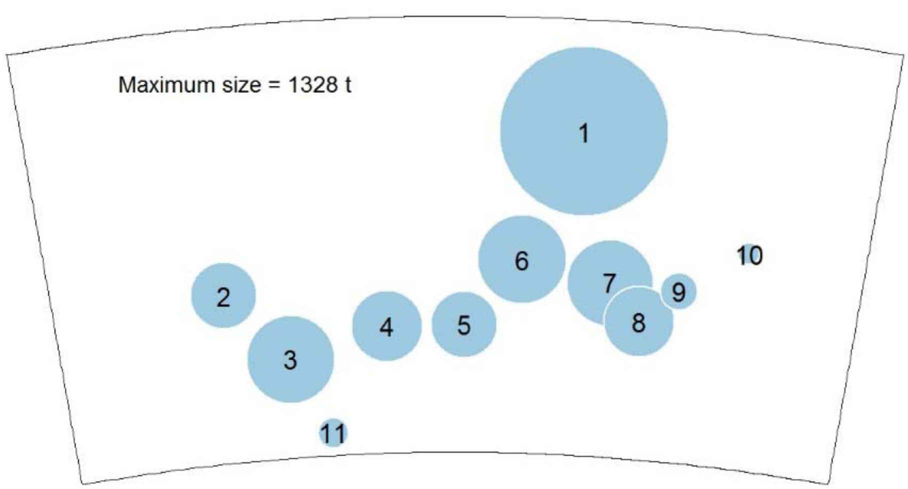

Figure 2: Spatial relationships of cumulative catch for each of the11 distinct seamount features (numbered 1-11) in SSRU 88.2H (Figure 2a from Parker (2014)).

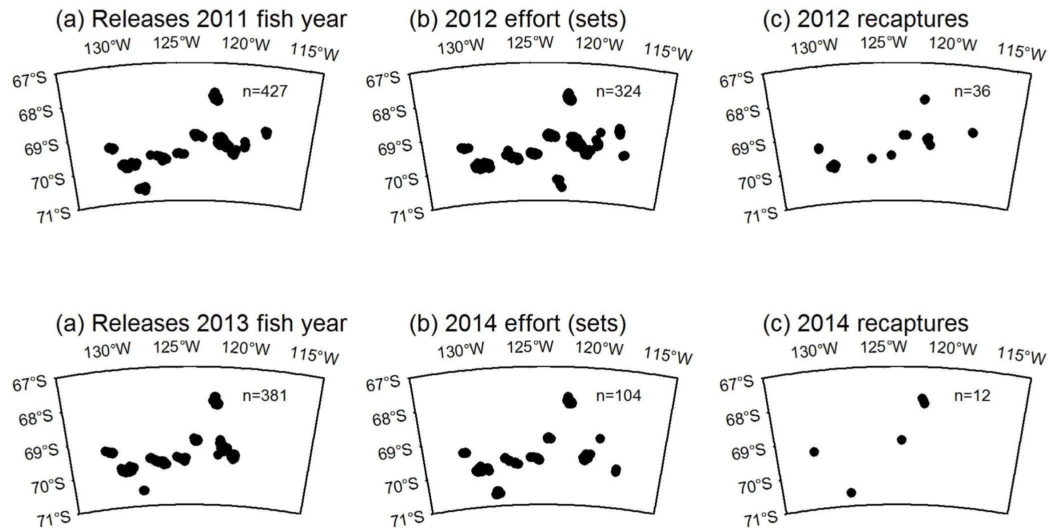

Figure 3: Case area SSRU 88.2H with all data used in this analysis: the location of release of tagged fish (panels a), the subsequent location of effort (panels b) and recaptures of tagged fish (panels c) for the 2012 fishing season (top) and 2014 fishing season (bottom)

4. RESULTS

Bayesian diagnostics such as visual inspection of trace plot and Geweke's Diagnostics (Geweke, 1992) were satisfied: Summary distributions were derived by calculating the Tag Spatial Overlap statistic (Equation 2.1) and the Tag Spatial Bias statistic (Equation 2.4) for each sample of the MCMC chain. The figures are histograms of these statistics that represent the marginal posterior for each statistic. The marginal posterior for the tag overlap statistic (left panel) and tag spatial bias for the two time periods show a shift in the distributions of both statistic between years (**Error! Reference source not found.**).

The results vary between years: 2012 has a TSO statistic of 70% (95% CIs 53–80%) and a TSB statistic of 88% (95% Cis 64–114%). The 2014 data has a TSO statistic of 60% (95% CIs 50-69%), and a TSB statistic 192% (95% CIs 141-268%).

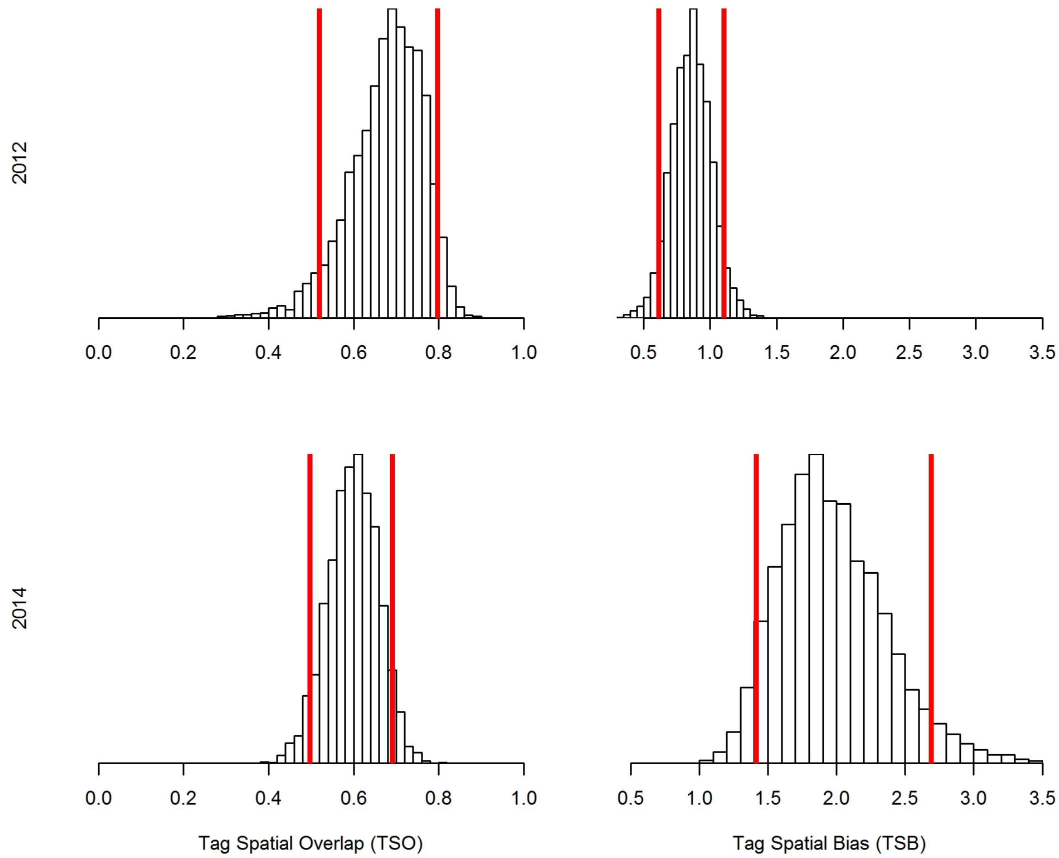

Figure 4: Marginal posterior distributions of both the TSO and TSB statistics for the 2012 (top) and 2014 (bottom) test cases. Vertical red lines represent the 95% credible intervals.

5. DISCUSSION

How much of the biomass in an area is accounted for using mark recapture observations when there may be poor spatial overlap between cells where tagging has occurred and where subsequent effort takes place? Understanding the spatial coverage of the release of tagged fish and the subsequent spatial coverage of the recovery fishing effort is an important consideration when interpreting mark-recapture data. We have developed a Tag Spatial Overlap (TSO) statistic and a Tag Spatial Bias (TSB) statistic to capture these processes.

The TSO statistic represents the percentage of estimated biomass that is informed by mark-recapture observations. This statistic behaved as expected, that is expressing a decrease from 2012 to 2014. Although the lower 95% credible intervals did not change much between the two year examples (approximately 3% difference). There was a decrease in the fiftieth percentile of 14% from 2012 to 2014, and a decrease in the 95% upper credible interval.

The TSB statistic represents the bias of mark-recapture biomass estimates when compared to the total estimated vulnerable biomass. This statistic suggested that the Chapman biomass estimate in area SSRU 88.2H in 2012 was between 64–114% of the estimate indexed by the combination of mark-recapture and catch rate data. One interpretation of this result, is that true vulnerable biomass could be as much as 36% lower or 14% higher than the Chapman biomass estimator alone implies. In 2012 the

TSB statistic was quite different with a value of 192% (95% CIs 141-268%), indicating the Chapman estimate of biomass was likely an over-estimate in that year.

The Tag Spatial Bias (TSB) statistic appeared sensitive to the relative weightings between observation data sets. Further investigation of the choice of likelihoods and the sensitivity of the estimators to data weighting are needed. In addition, it would be useful to expand the method to consider the sequential annual estimates as a time series within a single model. In this case study there were natural features (seamounts) that aided in dissecting a large area into smaller spatial cells. Additional work that would complement this paper would be a method to objectively scale and allocate data to spatial cells.

These statistics can be thought of as tools analogous to the tag-size overlap statistic tool in place since the 2011/12 season (Conservation Measure 41-01 Annex C, CCAMLR-XXIV, 2015; Ziegler, 2012). Potential uses for the TSO statistic can include creating a target spatial overlap level for research proposals which intend on developing a Chapman estimates of biomass, and diagnostics for mark-recapture biomass estimates. Potential uses for the TSB could include correcting for bias offsets in integrated assessments in the future, but the sensitivity of this statistic to the input distribution/assumptions suggests that further work is required to generate a more stable statistic before its application is considered.

6. ACKNOWLEDGEMENTS

The authors would like to thank the CCAMLR Secretariat, scientific observers and fishing company staff who collected the data used for this analysis. We would like to thank Steve Parker for review, and the New Zealand Antarctic Fisheries Working Group for helpful discussions and input into this paper. This project was funded by the National Institute of Water & Atmospheric Research Ltd. (NIWA) under NIWAs Fisheries Centre Research Programme.

7. REFERENCES

Burch, P., Parker, S. J., & Welsford, D. (2015). Quantifying uncertainty in the Chapman mark-recapture estimate of abundance (pp. 9). Hobart, Australia: CCAMLR.

CCAMLR-XXIV. (2015). *Schedule of Conservation Measures in force 2015/16 season.* Retrieved from Hobart, Australia:

Chapman, D. G. (1951). *Some properties of the hypergeometric distribution with applications to zoological sample censuses*. Berkeley: University of California Press.

Fournier, D., & Archibald, C. P. (1982). A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences, 39*, 1195–1207.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics, 4* (pp. 169–194). Oxford: Clarendon Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1).

Hillary, R. (2009). Assessment and tag program adaption methods for exploratory fisheries in the CAMLR Convention area: an example application for division 58.4.3a. *CCAMLR Science, 16*, 101-113.

Methot, R. D. (1990). Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *International North Pacific Fisheries Commission Bulletin, 50*, 259–277.

Mormede, S., Dunn, A., & Hanchet, S. M. (2014). A stock assessment model of Antarctic toothfish (*Dissostichus mawsoni*) in the Ross Sea region incorporating multi-year mark-recapture data. *CCAMLR Science, 2014*(21), 39-62.

Parker, S. J. (2014). Analysis of seamount-specific catch and tagging data in the Amundsen Sea, SSRU 88.2H (pp. 23). Hobart, Australia: CCAMLR.

SC-CAMLR-XXXIV. (2015, October 2015). *Report of the thirty-fourth meeting of the scientific committee.* Paper presented at the CCAMLR, Hobart, Australia.

Ziegler, P. E. (2012). Influence of the quality and quantity of data from a multi-year tagging program on bias and precision of biomass estimates from an integrated stock assessment - update (pp. 28). Hobart: CCAMLR.