



HAL
open science

Extracting abundance indices from longline surveys : method to account for hook competition and unbaited hooks.

Marie-Pierre Etienne, Shannon Obradovich, Lynne Yamanaka, Murdoch
Mcallister

► To cite this version:

Marie-Pierre Etienne, Shannon Obradovich, Lynne Yamanaka, Murdoch Mcallister. Extracting abundance indices from longline surveys: method to account for hook competition and unbaited hooks.. 2010. hal-00481079v1

HAL Id: hal-00481079

<https://hal.science/hal-00481079v1>

Preprint submitted on 5 May 2010 (v1), last revised 11 May 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extracting abundance indices from longline surveys: method to account for hook competition and unbaited hooks.

Marie-Pierre Etienne, Shannon Obradovich, Lynne Yamanaka, Murdoch McAllister

May 6, 2010

Abstract

The most commonly used relative abundance index in stock assessments of longline fisheries is catch per unit effort (CPUE), here defined as the number of fish of the targeted species caught per hook and minute of soak time. Longline CPUE can be affected by interspecific competition and the retrieval of unbaited or empty hooks, and interannual variation in these can lead to biases in the apparent abundance trends in the CPUE. Interspecific competition on longlines has been previously studied but the return of empty hooks is ignored in all current treatments of longline CPUE. In this work we propose some different methods to build indices to address the interspecific competition that relates to empty hooks. We show that in the absence of information about empty hooks, the relative abundance estimates have constant biases with respect to fish density and this is typically not problematic for stock assessment. The simple CPUE index behaves poorly in every scenario. Understanding the reasons for empty hooks allows selection of the appropriate index. A scientific longline survey is conducted every two years in the Strait of Georgia, British Columbia by Fisheries and Oceans Canada. The above methods are applied to build the time-series of indices from 2003 to 2009 for quillback rockfish (*Sebastes maliger*). Due to variation in the incidence of non-target species, the index trend obtained is moderately sensitive to the choice of the estimator.

Résumé

L'halieutique étudie les conséquences des politiques de gestion des pêches en modélisant l'abondance des populations d'intérêt. Des indices d'abondance sont utilisés pour ajuster des modèles de dynamique de population. Pour les pêches à la palangre l'indice d'abondance le plus couramment utilisé est le nombre de prises par unité d'effort (CPUE), c'est-à-dire le nombre de poissons de l'espèce étudiée divisé par le nombre d'hameçons. La compétition entre espèces et les hameçons remontés sans appât peuvent influencer ce type d'indice d'abondance. L'influence de la compétition entre espèces sur les indices d'abondance de palangre a déjà été étudiée

mais les hameçons remontés sans appât et sans prise sont systématiquement ignorés. Dans cet article nous proposons différentes méthodes pour construire des indices d'abondance pour la palangre qui tiennent compte de la compétition et des hameçons vides. Nous montrons que même en cas d'absence d'information sur l'origine des hameçons vides, les indices proposés ont un biais constant, indépendant du niveau de compétition ce qui ne pose pas de problème dans le cadre de la gestion de stock. Les CPUE conduisent à de mauvaises estimations quelque soit le niveau de compétition et la proportion d'hameçons vides. Comprendre la cause des hameçons vides permet de sélectionner l'indice le plus approprié. Une campagne de suivi scientifique est organisée tous les deux ans dans le détroit de Georgia en Colombie Britannique par Pêches et Océans Canada. Les méthodes présentées sont appliquées à ce cas d'étude pour construire des séries d'abondance de 2003 à 2009 pour des Sébastes de Roches *Sebastes maliger*.

1 Introduction

Many fish stock assessments derive information about stock trends from analyses of longline catch and effort records (IPHC, ICCAT, IATTC references). The classical relative abundance index, Catch Per Unit Effort (CPUE), for populations monitored using longline gear, is defined as the average number of individuals of target species caught per hook and minute of soak time. This commonly used CPUE index of abundance ignores the variability introduced by the competition for baited hooks within and between species. Somerton and Kikkawa (1995) propose two different indices, based on instantaneous rate of catch from longline survey records which takes interspecific competition into account. Haimovic and Avila-da Silva (2007) claim that CPUE and the instantaneous rate give the same results. Another important source of variability in the abundance indices which has received relatively little attention is the presence of empty hooks i.e., hooks returning without bait or fish.

This paper generalizes the use of instantaneous rates of catch from longline surveys as relative abundance indices (as proposed by Somerton and Kikkawa (1995) and Rothschild (1967)) and evaluates alternative approaches to dealing with empty hooks in the formulation of longline survey stock trend indices.

The paper is divided into three main sections after the introduction. Section 2 reviews existing methods for deriving relative abundance indices from longline catch and effort records. Section 3 presents our generalization of the previous methods to account for empty hooks. Some simulation studies are conducted to compare the indices under different levels of interspecific competition and sources for empty hooks. Section 4 gives the results of the simulation studies and illustrates the behaviour of the indices for monitoring the abundance of quillback rockfish (*Sebastes maliger*). The full technical details are outlined in the appendix.

2 Review of methods to derive abundance indices from longline records

2.1 Catch Per Unit Effort (CPUE)

In a longline survey CPUE is defined as the ratio between the number of fish of the target species caught N_T and the number of hooks N times the soak time S :

$$CPUE = \frac{\#Target}{\#hooks \times SoakTime} = \frac{N_T}{N S}.$$

This index ignores the effects of competition and gear saturation.

2.2 The simple exponential model

Somerton and Kikkawa (1995) have proposed two alternative approaches to deal with the issues of hook competition and gear saturation. Only one is useful in our context because dealing with gear saturation requires observed capture times which are not recorded in the available dataset. The number of available baits on the longline is supposed to decrease at an exponential rate measuring the overall pressure on the hooks. This overall pressure may be split into a sum of relative abundance indices per species. Using this approach the k^{th} catch $C_{r,k}$ for a soak time S and for species $r = 1, \dots, R$ is given by:

$$C_{r,k} = \frac{\lambda_r}{\lambda} N (1 - e^{-\lambda S}) + \varepsilon_{r,k}, \quad \text{with } \varepsilon_{r,k} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \quad (1)$$

where N stands for the initial number of hooks, λ is the overall pressure on hooks, and $\lambda = \sum_r \lambda_r$ is the sum of the specific abundance index per species, and k is the number of the set. The parameters $(\lambda_1, \dots, \lambda_R)$ defined a relative abundance index for each t species¹.

A few of the drawbacks of this Simple Exponential Model (SEM) are the assumptions of normality and homoscedasticity of the error terms. It is intuitive that the variability should be higher for species with higher relative abundance, i.e., the variance of the error term should not be constant but should depend in some way on λ_t . Furthermore $N_{t,k}$ is a discrete number (number of fish caught), potentially small and the normal assumption is not accurate in this case.

2.3 Multinomial Exponential Model

This section reviews the underlying ideas of Somerton and Kikkawa (1995) and proposes an alternative model which more closely mimics the behavior of the

¹ Mostly one particular species is the target species, let $r = 1$ while the other species, $r > 1$ are non-target species. In this context it is easier to consider only $\lambda_1 = \lambda_T$ the relative abundance for the target species and $\lambda_2 = \lambda_{NT}$ the relative abundance index which summarizes all the other species.

fish. This model has been originally proposed by Rothschild (1967) and describes how the catch of a target species could be reduced by the catch of other species.

Let us define T_T as the time it took to catch an individual from the target species on one particular hook. T_T is assumed to follow an exponential distribution of rate λ_T , i.e

$$\mathbb{P}(T_T \geq u) = e^{-\lambda_T u}.$$

T_{NT} is an exponential random variable with parameter λ_{NT} and models the time it took to catch an individual from any of the non-target species. We can define $T = \min\{T_T, T_{NT}\}$ as the time it took to catch an individual whatever the species it is. Thanks to the property of the exponential distribution, T is exponentially distributed with rate $\lambda = \lambda_T + \lambda_{NT}$. This property justifies the decomposition of the overall relative abundance as a sum of specific abundance given by Somerton and Kikkawa (1995).

After the soak period S , for one hook, there are only three possible outcomes:

- $\{I = 0\} = \{\text{The hook is still baited.}\}$. It means that the time for a capture is greater than the soak time. This event occurs with probability

$$\mathbb{P}(I = 0) = \mathbb{P}(T > S) = e^{-\lambda S}$$

- $\{I \neq 0\} = \{\text{The hook is no longer baited.}\}$.

$$\mathbb{P}(I \neq 0) = \mathbb{P}(T < S) = 1 - e^{-\lambda S}.$$

Given the hook is no longer baited, there are two possible outcomes: the catch is either from the target species, i.e. $\{I = T\}$, which occurs with probability

$$\mathbb{P}(I = T) = \mathbb{P}(T_T < T_{NT} | T < S) \mathbb{P}(T < S) = \frac{\lambda_T}{\lambda} (1 - e^{-\lambda S}),$$

or the catch is from a non-target species, corresponding to the event $\{I = NT\}$ which occurs with probability:

$$\mathbb{P}(I = NT) = \mathbb{P}(T_{NT} < T_T | T < S) \mathbb{P}(T < S) = \frac{\lambda_{NT}}{\lambda} (1 - e^{-\lambda S}).$$

Assuming that all the hooks on a longline behave independently, the likelihood is given by

$$L(\lambda_T, \lambda_{NT}) = \binom{N}{N_B} \binom{N_T}{N_T + N_{NT}} (e^{-\lambda S})^{N_B} (1 - e^{-\lambda S})^{N_T + N_{NT}} \left(\frac{\lambda_T}{\lambda}\right)^{N_T} \left(\frac{\lambda_{NT}}{\lambda}\right)^{N_{NT}}, \quad (2)$$

where

- N is the number of hooks on the longline,
- N_B is the number of baited hooks at the end of the soak time,
- N_T is the number of individuals of the target species caught,
- N_{NT} is the number of individuals of the non-target species caught.

The combinatorial terms arise since all the hooks are considered independent and the order of the catch on the longline has no importance. This model was originally proposed by Rothschild (1967) although presented here with a slightly different approach. This model is called Multinomial Exponential Model (MEM) since the vector (N_B, N_T, N_{NT}) follows a multinomial distribution whose vector of probability depends on an exponential term.

If λ_{NT} is larger than λ_T , it corresponds to a high level of competition: for a given relative abundance of the target species λ_T , the catch decreases as λ_{NT} , the non-target species relative abundance, increases.

2.4 Links between the indices

2.4.1 Links between MEM and SEM

The expected number of fish caught of the target species N_T is the same under the MEM and SEM assumptions and is given by:

$$\mathbb{E}(N_T) = N \frac{\lambda_T}{\lambda} (1 - e^{-\lambda S}).$$

Moreover, the models share the same parameters, λ_T and λ_{NT} . The main difference is the error term. In the SEM, the error term is normally distributed with a variance given by:

$$\text{Var}_{SEM}(N_T) = \text{Var}(N_{NT}) = \sigma^2,$$

while in the MEM the total number of fish caught has a multinomial distribution and the variance is given by:

$$\begin{cases} \text{Var}_{MEM}(N_T) = N \frac{\lambda_T}{\lambda} (1 - e^{-\lambda S}) \left(1 - \frac{\lambda_T}{\lambda} (1 - e^{-\lambda S})\right) \\ \text{Var}_{MEM}(N_{NT}) = N \frac{\lambda_{NT}}{\lambda} (1 - e^{-\lambda S}) \left(1 - \frac{\lambda_{NT}}{\lambda} (1 - e^{-\lambda S})\right). \end{cases}$$

Furthermore N_T and N_{NT} are assumed to be independent in the Simple Exponential Model and not in the Multinomial Exponential Model.

2.4.2 Links between CPUE and MEM

Under the MEM assumption, the expected CPUE of the target species is given by:

$$\mathbb{E}(CPUE) = \frac{\lambda_T}{\lambda S} (1 - e^{-\lambda S}) \xrightarrow{\lambda S \rightarrow 0} \lambda_T$$

If the overall density index λ is small enough, meaning that there is little competition, CPUE and the MEM index give the same results. This theoretical result is consistent with the expected behavior.

3 Dealing with empty hooks

In longline experiments, it is common for some hooks to return empty; the hook is no longer baited, but there is no fish on it. There could be several explanations for these empty hooks such as mechanical removal of bait during gear setting/retrieval, depredation by invertebrate or escape of fish with bait or removal of fish by predators.

In this paper, we consider the hypothesis that unbaited hooks arise only from the escape of fish. Therefore, the question about empty hooks is reduced to “How should the empty hooks be allocated to the different species?” These empty hooks provide information that we could use to improve the quality of our abundance indices. This section details modification of the MEM to incorporate this information and details some statistical properties of the indices built using those versions of MEM. It also describes different ways to include the empty hooks information into the SEM.

3.1 Full version of the Multinomial Exponential Model

We propose a modified version of the Multinomial Exponential Model to account for empty hooks. As opposed to the previous version of the MEM, here each fish caught has a probability of escaping equal to p_T , for target species, and p_{NT} , for non-target species.

We use three additional variables to fully specify the model: N_E is the number of observed empty hooks; $N_E^{(T)}$ (respectively $N_E^{(NT)}$) stands for the number of target species (respectively non-target species) individuals which have escaped and these two random variables are not observed. Assuming, as for the simple version of MEM, that all hooks are independent, we are able to conditionally describe the outcomes (figure 1).

- The number N_B of baited hooks retrieved at the end of the soak time is the realisation of a binomial random variable with probability of success $e^{-\lambda S}$.

$$N_B \sim \mathcal{B}(N, e^{-\lambda S}).$$

- Among the $N - N_B$ unbaited hooks, the total number of individuals from target species caught is $N_T + N_E^{(T)}$ and is also binomially distributed:

$$N_T + N_E^{(T)} | N_B \sim \mathcal{B} \left(N - N_B, \frac{\lambda_1}{\lambda} \right).$$

- Given $N_T + N_E^{(T)}$, the total number of individuals from target species on the longlines is N_T and is also binomially distributed:

$$N_T | N_T + N_E^{(T)} \sim \mathcal{B} \left(N_T + N_E^{(T)}, (1 - p_T) \right).$$

- Given $N_{NT} + N_E^{(NT)}$, the total number of individuals from non-target species on the longlines is N_{NT} and has also a binomial distribution:

$$N_{NT} | N_{NT} + N_E^{(NT)} \sim \mathcal{B} \left(N_{NT} + N_E^{(NT)}, (1 - p_{NT}) \right).$$

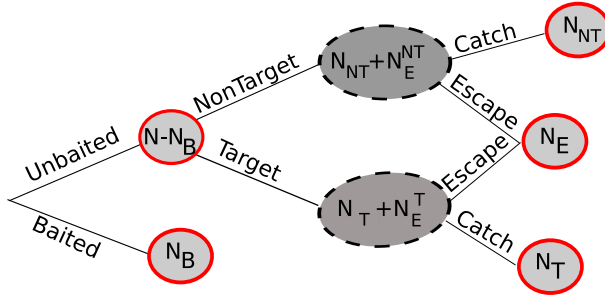


Figure 1: Conditional description of the model. The observed quantities are solid lines, the hidden quantities are dashed lines.

The full version of Multinomial Exponential Model is summed up through a probability tree in figure 1. $N_E^{(NT)}$ and $N_E^{(T)}$ are missing quantities but the sum N_E of these two quantities is observed. Annex B.1 gives the main step to define the likelihood of this model:

$$l(\lambda_T, \lambda_{NT}, p_T, p_{NT}) = \frac{N!}{N_B! N_T! N_{NT}! N_E!} (e^{-\lambda_S})^{N_B} (1 - e^{-\lambda_S})^{N - N_B} \left(\frac{\lambda_T}{\lambda} (1 - p_T) \right)^{N_T} \left(\frac{\lambda_{NT}}{\lambda} (1 - p_{NT}) \right)^{N_{NT}} \left(\frac{\lambda_T p_T + \lambda_{NT} p_{NT}}{\lambda} \right)^{N_E}$$

The full version of Multinomial Exponential Model may be considered as a multinomial distribution:

$$(N_B, N_T, N_{NT}, N_E) \sim \mathcal{M}(N, \boldsymbol{\alpha})$$

$$\text{with } \boldsymbol{\alpha} = \left(e^{-\lambda_S}, (1 - e^{-\lambda_S}) \frac{\lambda_T}{\lambda} (1 - p_T), (1 - e^{-\lambda_S}) \frac{\lambda_{NT}}{\lambda} (1 - p_{NT}), (1 - e^{-\lambda_S}) \frac{\lambda_T p_T + \lambda_{NT} p_{NT}}{\lambda} \right)$$

In this full version the model is not identifiable since an equivalent version could be expressed with only three parameters λ , $\lambda_T(1-p_T)$ and $\lambda_{NT}(1-p_{NT})$ (see annex B.2 for more details). Some additional information is required to estimate the parameters in this model. It is possible to add some biological knowledge on the probability of escape through prior distribution in a Bayesian framework but no information of this kind is available for our case study. This point is discussed in section 5. In a frequentist approach some particular solutions have to be chosen. In this paper we focus on two reasonable choices:

1. MEM1: empty hooks come only from non-target species, so p_T is assumed to be 0. Most of the time the target species is less abundant than all the non-target species. Allocating the empty hooks to the non-target species will at worst lead to an underestimation of the target species. Furthermore hooks and baits are selected to catch and retain the target species.
2. MEM2: another reasonable choice is to assume that the probability of escape are the same for target and non-target species, i.e $p_T = p_{NT}$. An empty hook has been bitten by a fish but no information about the species of this fish is available so that the empty hooks are allocated according to the relative densities of each group.

3.2 Maximum likelihood estimation of MEM

As all the longline sets are supposed to be independent, the complete likelihood is simply the product of the likelihood for each experiment given by formula 3. At this stage we have to consider two different situations.

If the soak times are different for all the longline sets, no analytical formula for the estimators is available, the estimation step has to be performed using a non linear optimization algorithm.

Mostly, the longline experiments have been designed to share the same soak time to reduce the causes of variation in the experiment. In this case, analytical formula can be derived for MEM1 and MEM2 because all the information can be summed up through the vector $(N_{B+}, N_{T+}, N_{NT+}, N_{E+})$ which corresponds to $(\sum_{l=1}^L N_{B_l}, \sum_{l=1}^L N_{T_l}, \sum_{l=1}^L N_{NT_l}, \sum_{l=1}^L N_{E_l})$, where l is the number of the longline set.

Even if the design of the experiment prescribes a constant soak time for each set, the actual soak time can differ slightly due to weather conditions or practical reasons. If the difference is not important, it is judicious to consider an single soak time (the mean for instance) to avoid the need of a numerical optimization which produces some instability in the estimations.

As detailed in appendix B.3, the maximum likelihood estimators are given

by:

$$\left\{ \begin{array}{l} \hat{\lambda}_T = \frac{N_{T+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\lambda}_{NT} = \frac{N_{NT+} + N_{E+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{p}_{NT} = \frac{N_{E+}}{N_{E+} + N_{NT+}}, \quad \hat{p}_T = 0 \end{array} \right. \quad \left\{ \begin{array}{l} \hat{\lambda}_T = \frac{N_{T+}}{N_{T+} + N_{NT+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\lambda}_{NT} = \frac{N_{NT+}}{N_{T+} + N_{NT+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{p}_T = \frac{N_{E+}}{N_{E+} + N_{T+} + N_{NT+}} = \hat{p}_{NT} \end{array} \right. \quad (4)$$

Maximum likelihood estimators are asymptotically unbiased and the covariance matrix is the inverse of Fisher Information Matrix (Severini, 2000). If the total number of hooks N is large enough, the joint distributions of these estimators can be approximated by a Multivariate Normal distribution. Asymptotically the covariance matrix for MEM1 is given by:

$$Cov_{MEM1} = \frac{\lambda_T \lambda_{NT}}{N(1 - e^{-\lambda S})} \begin{pmatrix} \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} \frac{\lambda_T}{\lambda_{NT}} + 1 & \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} - 1 & 0 \\ \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} - 1 & \frac{1-e^{-\lambda S}}{S^2 e^{-\lambda S} \lambda^2} \frac{\lambda_{NT}}{\lambda_T} + 1 & 0 \\ 0 & 0 & \frac{p_{NT}(1-p_{NT})\lambda}{\lambda_T \lambda_{NT}^2} \end{pmatrix}, \quad (5)$$

and for MEM2 by

$$Cov_{MEM2} = \frac{\lambda_T \lambda_{NT}}{N(1 - e^{-\lambda S})} \begin{pmatrix} \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} \frac{\lambda_T}{\lambda_{NT}} + \frac{1}{1-p} & \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} - \frac{1}{1-p} & 0 \\ \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} - \frac{1}{1-p} & \frac{(1-e^{-\lambda S})^2}{S^2 e^{-\lambda S} \lambda^2} \frac{\lambda_{NT}}{\lambda_T} + \frac{1}{1-p} & 0 \\ 0 & 0 & \frac{p(1-p)}{\lambda_T \lambda_{NT}} \end{pmatrix}. \quad (6)$$

The result proposed by Rothschild (1967) concerning the asymptotic variance for the simple version of MEM with $p_T = p_{NT} = 0$ should be the same replacing p or p_{NT} by 0 in the above formula. Nevertheless the two formulas are not compatible even if the estimators are. We suspect a mistake in the formula proposed by Rothschild (1967).

3.2.1 Bayesian framework for MEM

The multinomial exponential model could also be estimated in a Bayesian framework. The full specification of the Bayesian version of the MEM, requires the definition of some prior distributions for the parameters $(\lambda_T, \lambda_{NT}, p_T, p_{NT})$. The relative abundance index is always less than 1. Therefore, in our study, the priors have been chosen as poorly informative and independent.

$$\lambda_T \sim \beta(1, 1) \quad \lambda_{NT} \sim \beta(1, 1).$$

If there is no informative prior on the probability of escape, the model is still non-identifiable. In a Bayesian framework, an identifiable model can be diagnosed since the posterior distribution is the same than the prior distribution. Nothing has been learned from the data.

To avoid this problem, some informative prior distributions could be defined using biological knowledge or field experiment. This aspect hasn't been investigated in this work. We just use the specific forms of the model (MEM1 or MEM2) to remove the problem of identifiability.

The estimation procedure has been implemented using the JAGS software (Plummer, 2008). An example of JAGS code is provided in section C.1 for MEM1 and C.2 for MEM2 and all other codes are available on request.

3.3 Estimation for other indices

This section will discuss the estimation steps of all the indices previously presented and highlights the key points of the procedure.

3.3.1 Multiple CPUE

In a given region, you could have several sets of longlines deployed, so the estimation step requires the fitting of one index using several observations.

If only one set is deployed, the CPUE is obtained by the catch over the soak time times the number of hooks. A simple generalization of this index is detailed in A and proposes to compute CPUE as

$$CPUE = \frac{\sum_{l=1}^L N_{T_l}}{\sum_l S_l N_l},$$

where $l = 1, \dots, L$ stands for the longline. If the number of hooks or the soak times are different, the generalized CPUE computed is a weighted average of all individual CPUE. A simple average is not satisfying, since for example, the average will attribute the same weight to an experiment with 200 hooks and to another experiment with only 50 hooks.

3.3.2 Simple Exponential model

In the following, two versions of SEM are derived depending on how unbaited hooks are considered:

1. SEM1: empty hooks are assumed to arise only from the non-target species and N_E and N_{NT} are pooled together.
2. SEM2: empty hooks are considered as a "third" species, and an additional relative abundance index is defined λ_E .

Hovgård and Lassen (2000) proposes to estimate λ using

$$\hat{\lambda}_{Hov} = \frac{-\log(N_B/N)}{S}.$$

If all longline sets share the same soak time and the same initial number of hooks N , the MLE for this model are almost the same than for the MEM

except concerning the empty hooks.

$$\begin{aligned}
SEM1 & \left\{ \begin{aligned} \hat{\lambda}_T &= \frac{N_{T+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\lambda}_{NT} &= \frac{N_{NT+} + N_{E+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\sigma}^2 &= \frac{1}{2L} \sum_{l=1}^L \left(N_{T_l} - N \frac{\hat{\lambda}_T}{\lambda} (1 - e^{-\hat{\lambda}S}) \right)^2 + \left(N_{NT_l} + N_{E_l} - N \frac{\hat{\lambda}_{NT}}{\lambda} (1 - e^{-\hat{\lambda}S}) \right)^2 \end{aligned} \right. \\
SEM2 & \left\{ \begin{aligned} \hat{\lambda}_T &= \frac{N_{T+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\lambda}_{NT} &= \frac{N_{NT+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\lambda}_E &= \frac{N_{E+}}{N_+ - N_{B+}} \frac{1}{S} \log \left(\frac{N_+}{N_{B+}} \right) \\ \hat{\sigma}^2 &= \frac{1}{3L} \sum_{l=1}^L \left\{ \left(N_{T_l} - N \frac{\hat{\lambda}_T}{\lambda} (1 - e^{-\hat{\lambda}S}) \right)^2 + \left(N_{NT_l} - N \frac{\hat{\lambda}_{NT}}{\lambda} (1 - e^{-\hat{\lambda}S}) \right)^2 + \right. \\ & \quad \left. \left(N_{E_l} - N \frac{\hat{\lambda}_E}{\lambda} (1 - e^{-\hat{\lambda}S}) \right)^2 \right\} \end{aligned} \right.
\end{aligned}$$

To estimate σ at least two longline sets are required: otherwise those estimations correspond to perfect match and there is no additional variability. Since we are only interested in building abundance indices for the target species and since the estimations of $\hat{\lambda}_T$ in SEM1 and SEM2 are the same, in the following we only call this index SEM.

When the soak times or the initial number of hooks are different, a numerical optimisation algorithm has to be used to defined the MLE for SEM1 and SEM2. This approach should be avoided if possible due to numerical instability. From a practical point of view the `nlm` function available in R software (R Development Core Team, 2009) behaves badly. In this work we directly optimize the loglikelihood function using function `optim`.

The analytical formulas for λ_T are exactly the same for SEM and MEM1: if soak times and initial number of hooks are the same for all sets, there is absolutely no difference between these two indices concerning the relative abundance of the target species, which is not true for the non-target species. But this analytical formula is only valid for SEM when soak times and initial number of hooks are the same while MEM only requires the same soak times.

3.4 Simulation studies

Bias and variability of the estimators are evaluated in this section through simulation studies. Several plausible scenarios have been studied to give some robust results and advice about the behaviour of all the indices.

Specifying data generators is the first step of simulation studies. In our specific case, we have at first sight different solutions. We can use the Simple Exponential Model but this model doesn't simulate any empty hooks, and it provides some non-integer catch because of the normal hypothesis. The two other solutions (ie MEM1 and MEM2) will be used to study the different scenarios.

This choice of data simulator gives obviously an advantage to MEM1 if the data were simulated with MEM1 and to MEM2 in the other case.

For one set of fixed parameters (that is λ_T , λ_{NT} , L the total number of sets, S the soak time, N the number of hooks on a longline) 5000 fake datasets are generated and the corresponding estimated values for λ_T and λ_{NT} have been computed. A relative bias and a coefficient of variation are derived from these simulations. To study the impact of the estimation via numerical algorithm, we also compute the estimators by maximizing the log likelihood using a non linear optimization algorithm.

The values λ_T and λ_{NT} need to be chosen to reflect a plausible situation. In this work, four values of each parameter have been used and all of the sixteen combinations of these two parameters have been addressed.

Three different scenarios for unbaited hooks have been simulated:

- Sc.1) There are no empty hooks. Each unbaited hook has caught a fish. This situations corresponds to $p_{NT} = p_T = 0$.
- Sc.2) The ability to escape is the same the across the species. The probability of escape is set to 20%, this corresponds to $p_T = p_{NT} = 0.2$.
- Sc.3) non-target species individuals are better at escaping. The probability of escape is set to $p_{NT} = 0.2$ for the non-target species and to $p_T = 0.02$ for the target species.

The results of this simulation study are presented in section 4.1.

3.5 Description of the B.C. inshore rockfish longline survey

Since 2003, Fisheries and Oceans Canada has conducted an annual scientific longline survey in the Strait of Georgia with the fisheries research vessel CCGS Neocaligus. Different regions of the Strait are covered each year, resulting in each statistical area (PMFSC) being surveyed every two to three years. A 2 km by 2 km grid is overlaid on all inshore rockfish habitat up to 100 m in depth, as determined using Canadian Hydrographic Service (CHS) charts. These blocks are stratified into shallow (41-70 m) and deep (71-100m) and 8% of the blocks, in a given statistical area, are randomly selected for fishing each year (see Lochead and Yamanaka 2007 for further details)..

The snap-type longline gear consists of 1800 ft of leaded groundline with 225 circle hooks (13/0) spaced 6 to 12 ft apart. Each hook is attached to the snap by a 1.2 ft perlon gangion, crimped at both ends, and with a swivel at the hook. The hooks are baited with Argentinean squid (sp?). Soak time for each set is approximately two hours, measured as the time from deployment of the last hook to retrieval of the first hook on board.

As the gear is retrieved, the condition of each hook is recorded as returning with bait, with catch on the hook, empty (i.e. without bait or catch on the hook), or unknown, if the hook does not return. Catch is recorded to the species level for both fish and invertebrates.

4 Results

4.1 Simulation studies

4.1.1 Competition but no empty hooks

Figure 2 presents the bias for two indices in absence of empty hooks. SEM, MEM1, MEM2 produce exactly the same results because there is no empty hooks and the analytical formula have been used.

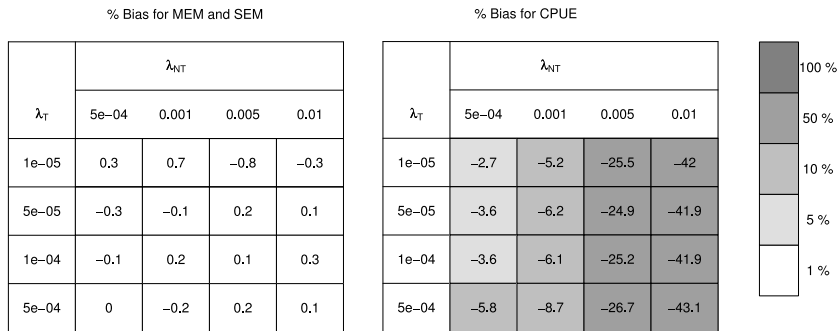


Figure 2: Relative bias, defined as the absolute value of the bias divided by the true value $|\lambda_T - \hat{\lambda}_T|/\lambda_T$, expressed in percentage computed over 5000 simulations for 220 hooks per longline and 20 sets. Bias for SEM1, SEM2, MEM1 and MEM2 is the same in this case. On the right the bias computed for CPUE indices show that the bias increases with the increase of relative abundance of the non-target species.

The estimations are unbiased for abundance indices built on the exponential model and so competition is effectively taken into account. On the other hand bias in the CPUE index increases with the increase of the relative density of non-target species. This result is a confirmation that CPUE index strongly depends on competition and should be avoided. This behavior is always the same in all situations which have been addressed in this simulation study.

The coefficient of variation presented in table 1 depends on the number of data relying on N and L but also on the relative abundance, parameters λ_T and λ_{NT} . This coefficient could be very high for low relative abundance situation.

The coefficient of variation of the estimators for the relative abundance decreases with the expected number caught, which is in this study $LN \frac{\lambda_T}{\lambda} (1 - e^{-\lambda S})$. The relation between these two quantities is illustrated by figure 3. This expected number caught describes the actual available information on λ_T .

The bias study shows that CPUE behaves really badly when interspecific competition occurs. In the following the results concerning this index will not be shown.

λ_T	λ_{NT}			
	5e-04	0.001	0.005	0.01
1e-05	43.2	44.8	50.9	56.7
5e-05	19.5	20.1	22.2	25.3
1e-04	13.8	14.4	15.9	17.9
5e-04	6.2	6.4	7.4	8.1

Table 1: Coefficient of variation (%) in estimates of λ_T using the MEM models. Results For SEM, MEM1 and MEM2 are exactly the same since the soak time is shared by all longline sets and there is no empty hook.

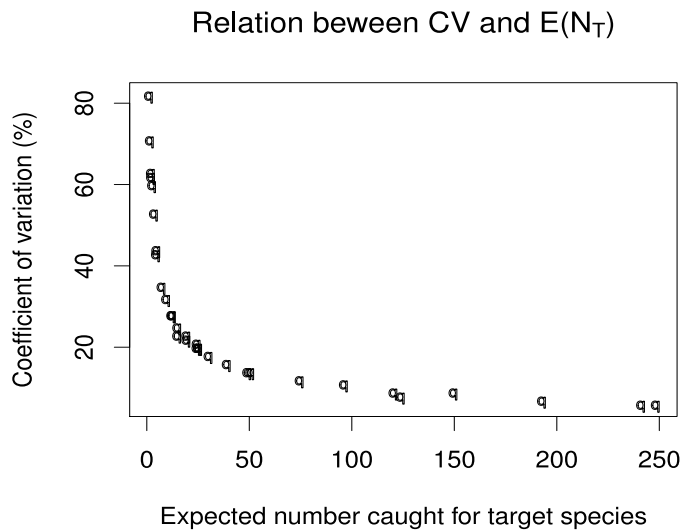


Figure 3: Coefficient of variation for the estimates of λ_T for MEM1 computed over 5000 simulations as a function of the expected number of catch for target species

4.1.2 Competition and empty hooks

SEM and MEM have the same behaviour when no empty hooks are present in the dataset. Simulations of empty hooks allow the comparison of the respective behavior of those indices in presence of unbaited hooks. In our simulation context, the soak time and the initial number of hooks are shared by all sets, so that analytical formula can be used to compute the indices and SEM and MEM1 produce the same results.

Since MEM1 and MEM2 relies on two different hypothesis concerning the origin of the empty hooks, they give some different results. When all species are as good at escaping (Sc.2) MEM2 produces unbiased estimators of the relative abundance while SEM1, SEM2 and MEM1 tend to underestimate this relative abundance. In the simulations used to produce figure 4, the probability of escape for non-target and target fish has been set to 20%, which corresponds to the underestimation of 20% for the relative abundance λ_T .

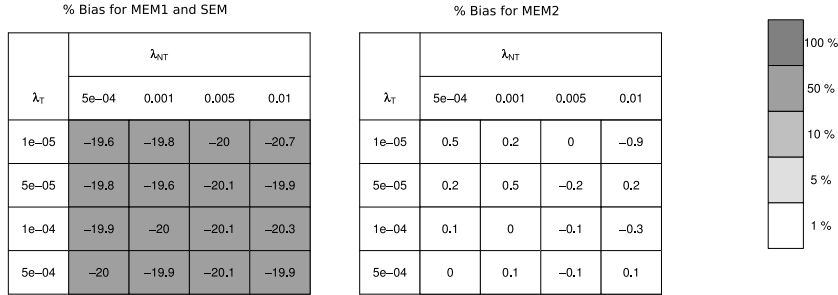


Figure 4: Simulations under Scii. MEM1, SEM1 and SEM2 produce the same results and tend to underestimate the relative abundance. MEM2 is the "true" model in this situation and tends to produce unbiased estimate.

When the non-target species are better at escaping, MEM2 tends to overestimate the relative abundance of the target species. Results presented in figure 5 are produced when $p_T = 0.02$ and $p_{NT} = 0.2$. MEM2 attributes a proportion of the empty hooks to the target species, this proportion depends on λ_T and λ_{NT} . Since in this simulation most of the empty hooks arise from the non-target species, the higher the relative abundance of non-target species is, the more the relative abundance of target species is overestimated. The bias in the estimates for MEM1 is constant and equals 2% which corresponds the missed fish from the empty hooks.

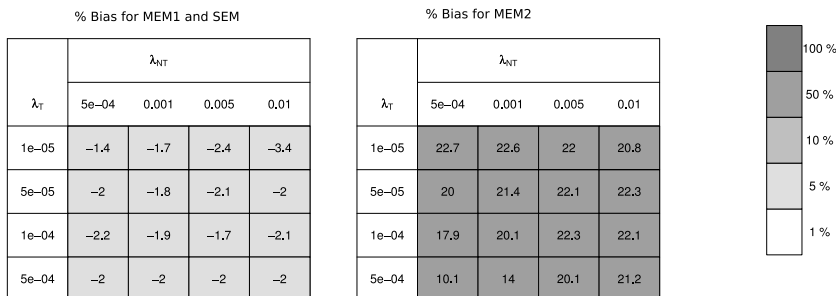


Figure 5: Simulations under Sc1: MEM1 and SEM underestimate the true relative abundance by exactly the probability of escape for non-target species. MEM2 highly overestimates the relative abundance.

4.1.3 Numerical instability

In order to study the numerical instability of the optimization algorithm, the maximum likelihood estimators have been computed through the analytical formula and using numerical optimisation algorithm on the same data set (with shared S and shared N). Whatever scenario is used, the optimisation algorithm behaves well, i.e, less than 5% of difference between the estimates computed using the analytical formula and the estimates obtained by numerical optimization except when the ratio between the relative abundance of non-target species λ_{NT} and the relative abundance of the target species λ_T is very high. In the extreme case, with $\lambda_{NT} = 0.01$ and $\lambda_T = 1e - 05$ the average difference between the two estimates varies from 10% to 40%. This bad behaviour occurs when the log-likelihood peak is not strong enough. Some examples of this poor behaviour concerning the optimisation step are illustrated on the real data in section 4.2.

4.2 Rockfish survey results

Figure 6 shows the different relative abundance indices obtained using the scientific survey described in section 3.5. The confidence intervals have been computed using a bootstrap procedure with 5000 resamples.

The estimate for the numerical version of SEM index exhibits a considerable difference in trends and the confidence interval associated to this estimate is huge. This is due to numerical instability problem. Indeed, there are very few quillback caught in 2007 in area 13 as suggested by the decrease in the trend for all other indices. Therefore, the numerical optimization procedure behaves badly which produce a bad estimate but also a large variability using bootstrap procedure.

Since favorable habitat for target species are mainly located in area 12 and 13 and since the management isn't at the area scale but globally, data from area

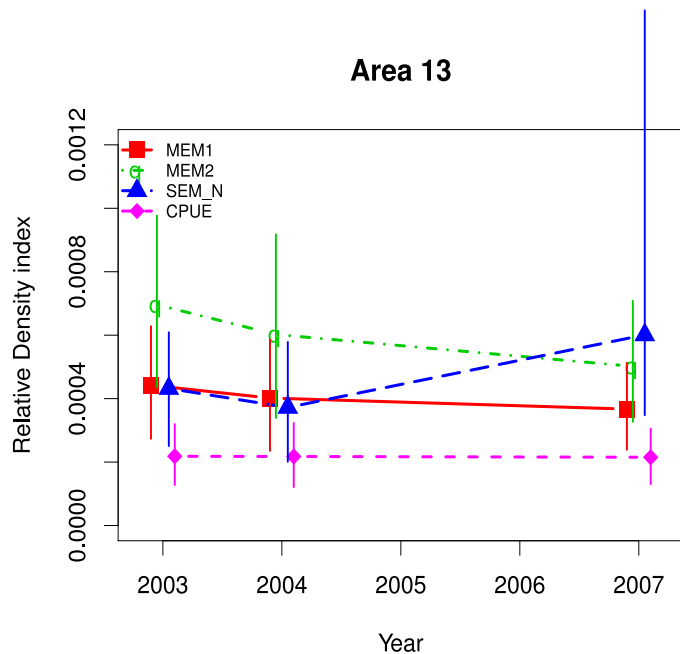


Figure 6: Four indices computed for area 13 for the quillback population. The numerical optimisation has some stability issues for year 2007, the results should be the same then MEM1 estimates.

12 and 13 have been pooled to form one dataset. The corresponding relative abundance time series is shown in figure 7

Pooling the data from both areas produce more precise estimates as shown in table 2 and avoids the numerical instability for the SEM index. The coefficient of variation is divided by almost two when using the whole dataset. We didn't include the other areas of the Strait of Georgia in the study since they are known to provide a poor habitat for quillback rockfish.

Given the uncertainty on the relative abundance indices, no change is statistically significant in the relative abundance of quillback population in the strait of Georgia.

5 Discussion

The first and most important conclusion of this study is that when competition between was accounted for simulated data, the classical longline CPUE estimators gave some strongly biased estimates of stock trends in all cases evaluated. In contrast, the exponential model-based estimators showed much less bias. Therefore it is advisable to avoid the use of classical CPUE as much as

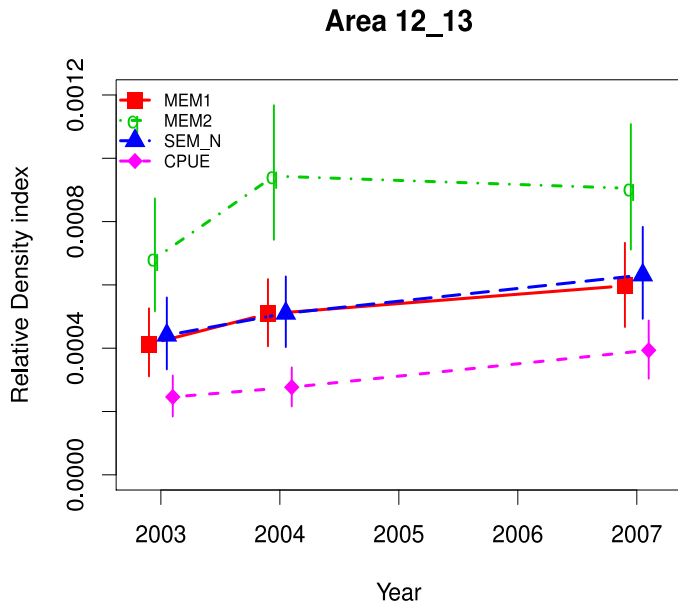


Figure 7: Four indices computed for area 12 and 13.

possible since this index doesn't take competition into account. From one year to another, the level of competition can vary and two CPUE indices computed during two different years are not comparable which is unacceptable for a relative abundance index.

In the absence of empty hooks, the abundance indices built on SEM or MEM are quite comparable even if the fundamental assumptions of SEM and MEM are very different. If N_T is considered as a sum over all hooks of the number of hooks which have caught a target individual, the central limit theorem claims that if the number of hooks is large enough N_T exhibits a normal distribution. The independence of hooks is not required, only some weak dependence is necessary (see Billingsley (1995) for central limit theorem under weak dependence conditions). SEM doesn't assume the independence on hooks but N_T and N_{NT} are considered as independent variables which is obviously false especially when this model is used to take competition into account. This explains how the number of non-target species caught N_{NT} could decrease the number of target species caught N_T . In contrast MEM models the dependence between N_T and N_{NT} but assumes the independence of hooks.

Ignoring empty hooks could produce bad indices because most of the empty hooks correspond to fish escapement which should be counted in the relative abundance index. Empty hooks are of major importance to building abundance indices even if there is no perfect solution to deal with them. Biological knowledge can be very useful for deciding how to deal with empty hooks and the Bayesian framework offers an intuitive way to use this kind of information to

Coefficient of variation (%)						
Index	Area 13			Area 12 and 13		
	2003	2004	2007	2003	2004	2007
MEM1	24.1	26.6	22.7	15.9	12.6	13.6
MEM2	23.5	29	23.1	15.6	13.6	13.3
SEM N	25.8	30.2	29.8	15.6	13.1	13.9
CPUE	26.4	28.1	24.7	16	13.4	14.1

Table 2: Coefficient of variation of the different relative abundance indices expressed in percent. The variability strongly decreases when data are pooled together. The large value for the coefficient of variation for 2007 for the numerical version of SEM is due to optimization instability.

remove the non identifiability problem. But this biological knowledge is hard to obtain since the escapement of fish from hooks is difficult to study. MEM1 and MEM2 require fully explicit choices concerning the empty hooks. Even if this required choice is hard to make, the explicit choice could still be discussed. In the SEM model the choice is made by default and it is even not explicit. We recommend that practitioners design studies to collect information about the ability of target and common non-target species to escape and work in a Bayesian framework using the code provided in Annex C. The major advantage is that the uncertainty about empty hooks is translated to uncertainty about the relative abundance indices. If absolutely no information is available, the recommendation is to use SEM1 since the bias on the relative abundance index doesn't depend on the abundance of non-target species. Furthermore, multiplicative constant bias as seen for MEM1 doesn't produce biased estimates of population dynamic parameters model since the multiplicative constant is absorbed by the coefficient of proportionality which links abundance indices and biomass.

One major drawback of all the models formulated in this paper is the assumption of constant relative abundance along the longline set which is obviously not true. Different approaches could be explored to avoid this assumption. The first one would be to record the change of habitat along the longline set and using this as a covariate in the model. But it is mostly not possible to have some information about the habitat except when the longline is observed during the soak with a Remotely Operated Vehicle. Another possibility would be to refine the modeling of the abundance index. We could for instance consider a local relative abundance index λ_{Th} at hook h defined as the sum of the main relative abundance λ_T plus a noise term. The noise term would be chosen as an autoregressive model for example to use the information of the hooks in the neighbourhood. This extra variability could account for the variability in the habitat. Another perspective for dealing with the variability along the longline is to define the abundance index as piecewise constant function along the longline and trying to detect the change in this function using the tools of change-point detection (Lavielle and Lebarbier, 2001).

The possible variation of λ_T during the soak time is also a question of interest. Some species could be more attracted by a fresh bait and therefore λ would be supposed to decrease with time. This question of attractivity of the baits has been studied by Fernö and Olsen (1995) in chapter 8 but currently there is no solution to take this into account when building some abundance indices.

References

- Billingsley, P. (1995). *Probability and Measure (Third ed.)*. John Wiley & sons.
- Fernö, A. and Olsen, S., editors (1995). *Marine Fish Behaviour in Capture and Abundance Estimation*. Wiley.
- Gelfand, A. and Smith, A. (1990). Sampling based approach to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.
- Haimovic, M. and Àvila-da Silva, A. O. (2007). A comparison of methods for estimating relative abundances in bottom longline fishery surveys. *Fisheries Research*, 85:226–231.
- Hovgård, H. and Lassen, H. (2000). *Manual on estimation of selectivity of gillnet and longline gears in abundance surveys*. FAO Fish. Tech. Paper.
- Lavielle, M. and Lebarbier, E. (2001). An application of mcmc methods for the multiple change-points problem. *Signal Processing*, 81(1):39–53.
- Plummer, M. (2008). *JAGS Version 1.0.2 Manual*. Lyon: International Agency for Research on Cancer.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rothschild, B. J. (1967). Competition for gear in a multiple-species fishery. *ICES Journal of Marine Science*, 31:102–110.
- Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford university Press, New York.
- Somerton, D. A. and Kikkawa, B. S. (1995). A stock survey technique using the time to capture individual fish on longlines. *Canadian Journal of Fisheries and Aquatic Science*, 52:260–267.

A CPUE computed from several observations

A CPUE index may be considered as the estimation of the probability of successes in a binomial trial. If N_T stands for the number of success and p is the probability of success over $N \times S$ trials, then

$$N_T \sim \mathcal{B}(p, N \times S).$$

The maximum likelihood estimator p is given by $N_T/(N \times S) = \text{CPUE}$. If independent records N_{T_i} for L longline sets are available with distribution $N_{T_i} \sim \mathcal{B}(p, N_i \times S_i)$, a sufficient statistic is the sum of all the catch N_T^+ which is distributed as

$$\sum_i N_{T_i} = N_T^+ \sim \mathcal{B}(p, \sum_i N_i \times S_i).$$

The maximum likelihood estimator is defined as

$$\frac{\sum_{i=1}^L N_{T_i}}{\sum_i S_i N_i},$$

which defines a generalized CPUE definition.

B More details about the MEM

B.1 Likelihood of MEM with empty hooks

This section gives the key steps of the calculus for the likelihood given in formula 3. We use brackets to denote *pdf*'s as many conditioning terms will appear in the probabilistic expressions derived from the full version of Multinomial Exponential Model. As in Gelfand and Smith (1990), the brackets denote either a density or a discrete probability distribution.

The likelihood is then defined by:

$$\begin{aligned} l(\lambda_T, \lambda_{NT}, p_T, p_{NT}) &= [N_B, N_T, N_{NT}, N_E | \lambda_T, \lambda_{NT}, p_T, p_{NT}], \\ &= [N_B | \lambda_T, \lambda_{NT}] [N_T, N_{NT}, N_E | N_B, \lambda_T, \lambda_{NT}, p_T, p_{NT}]. \end{aligned}$$

By definition of the model $[N_B | \lambda_T, \lambda_{NT}]$ is a binomial distribution. We need then to define the joint distribution of (N_T, N_{NT}, N_E) given N_B the total number of unbaited hooks. To obtain this distribution, we need to explicit the integration term over the hidden quantities $N_T^{(E)}$ and $N_{NT}^{(E)}$. To make the writing easier to follow all the parameters will be omitted in the conditioning term.

$$\begin{aligned}
[N_T, N_{NT}, N_E | N_B] &= \sum_{k=0}^{N_E} \left[N_T, N_{NT}, N_T^{(E)} = k, N_{NT}^{(E)} = N_E - k | N_B \right] \\
&= \sum_{k=0}^{N_E} \binom{N - N_B}{N_T + k} \left(\frac{\lambda_T}{\lambda} \right)^{N_T + k} \left(\frac{\lambda_{NT}}{\lambda} \right)^{N_{NT} + N_E - k} \binom{N_T + k}{k} p_T^k (1 - p_T)^{N_T} \\
&\quad \left(\frac{N_{NT} + N_E - k}{N_E - k} \right) p_{NT}^{N_E - k} (1 - p_{NT})^{N_{NT}} \\
&= \left\{ \left(\frac{\lambda_T}{\lambda} \right) (1 - p_T) \right\}^{N_T} \left\{ \left(\frac{\lambda_{NT}}{\lambda} \right) (1 - p_{NT}) \right\}^{N_{NT}} \frac{(N - N_B)!}{N_T! N_{NT}!} \\
&\quad \sum_{k=0}^{N_E} \frac{1}{(N_E - k)! k} \left(\frac{\lambda_T}{\lambda} p_T \right)^k \left(\frac{\lambda_{NT}}{\lambda} p_{NT} \right)^{N_E - k} \\
&= \frac{(N - N_B)!}{N_T! N_{NT}! N_E!} \left(\frac{\lambda_T}{\lambda} (1 - p_T) \right)^{N_T} \left(\frac{\lambda_{NT}}{\lambda} (1 - p_{NT}) \right)^{N_{NT}} \left(\frac{\lambda_T p_T + \lambda_{NT} p_{NT}}{\lambda} \right)^{N_E}
\end{aligned}$$

The likelihood is then obtained by combining the binomial distribution of N_B with this previous result to give equation 3.

B.2 Identifiability

To prove that the full version of the MEM is not identifiable it is sufficient to express the likelihood with only three parameters. Let us define $\alpha = \frac{\lambda_T}{\lambda} (1 - p_T)$ and $\beta = \frac{\lambda_{NT}}{\lambda} (1 - p_{NT})$. Therefore the likelihood given in equation 3 may be rewritten as:

$$l(\lambda, \alpha, \beta) = \frac{N!}{N_B! N_T! N_{NT}! N_E!} (e^{-\lambda S})^{N_B} (1 - e^{-\lambda S})^{N - N_B} (\alpha)^{N_T} (\beta)^{N_{NT}} (1 - \alpha - \beta)^{N_E}. \quad (7)$$

This form of the model is identifiable and it is called the regular form of the model.

B.3 Estimators

B.3.1 MEM Estimators

Using the regular form of the model given in equation 7, it is easy to derive the log likelihood and obtained the maximum likelihood estimators:

$$\hat{\lambda} = \frac{1}{S} \log \left(\frac{N}{N_B} \right), \quad \hat{\alpha} = \frac{N_T}{N_T + N_{NT} + N_E}, \quad \hat{\beta} = \frac{N_{NT}}{N_T + N_{NT} + N_E}.$$

Those estimators are quite intuitive: for example α parameter represent the relative density of catchable individuals from target species and it is naturally estimated as the ratio between the the catch of target species and the total

number of catch including empty hooks. The ambiguity of the model lies then in the definition of the relative density, how much the catchable density has to be increased?

The maximum of the loglikelihood function equals

$$\log \left(\frac{N!}{N_B! N_T! N_{NT}! N_{NE}} \right) + \log \left(\frac{N_B^{N_B} N_T^{N_T} N_{NT}^{N_{NT}} N_{NE}^{N_{NE}}}{N^N} \right)$$

This value could be used to compute an AIC criteria.

In the following we will derive the maximum likelihood estimators in our two special cases of interest.

B.3.2 MEM1 Estimators

MEM1 corresponds to the assumption that the target species can't escape, so that $p_T = 0$, in this context the log likelihood is given by

$$L(\lambda_T, \lambda_{NT}, p_{NT}) = K - N_B S \lambda + (N - N_B) \log(1 - e^{-\lambda S}) + N_T \log(\lambda_T) \\ + (N_{NT} + N_E) \log(\lambda_{NT}) + N_{NT} \log(1 - p_{NT}) - (N - N_B) \log(\lambda) + N_E \log(p_{NT})$$

The first derivatives are then derived:

$$\frac{\partial L}{\partial \lambda_T} = -N_B S + \frac{(N - N_B) S e^{-\lambda S}}{1 - e^{-\lambda S}} + \frac{N_T}{\lambda_T} + \frac{N - N_B}{\lambda}, \\ \frac{\partial L}{\partial \lambda_{NT}} = -N_B S + \frac{(N - N_B) S e^{-\lambda S}}{1 - e^{-\lambda S}} + \frac{N_{NT} + N_E}{\lambda_{NT}} + \frac{N - N_B}{\lambda}, \\ \frac{\partial L}{\partial p_{NT}} = -\frac{N_{NT}}{1 - p_{NT}} + \frac{N_E}{p_{NT}}.$$

The maximum likelihood estimators given on the left side of equation 4 are then obtained by determining the roots of this set of equations.

If N the number of hooks is large enough, the distribution of the estimators may be approximated by a normal distribution with mean $(\lambda_T, \lambda_{NT}, p_{NT})$ and a matrix variance which is the inverse of the Fisher information matrix F defined by

$$F = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \right],$$

where θ represents a generic vector of parameters.

It can be shown that the asymptotic variance matrix for MEM1 is given by:

$$Cov_{MEM1} = \frac{1}{N} \begin{pmatrix} \frac{\lambda_T \lambda_{NT}}{1 - e^{-\lambda S}} + \frac{1 - e^{-\lambda S}}{S^2} \frac{\lambda_T^2}{e^{-\lambda S} \lambda^2} & -\frac{\lambda_T \lambda_{NT}}{1 - e^{-\lambda S}} + \frac{1 - e^{-\lambda S}}{S^2} \frac{\lambda_T \lambda_{NT}}{e^{-\lambda S} \lambda^2} & 0 \\ -\frac{\lambda_T \lambda_{NT}}{1 - e^{-\lambda S}} + \frac{1 - e^{-\lambda S}}{S^2} \frac{\lambda_T \lambda_{NT}}{e^{-\lambda S} \lambda^2} & \frac{\lambda_T \lambda_{NT}}{1 - e^{-\lambda S}} + \frac{1 - e^{-\lambda S}}{S^2} \frac{\lambda_{NT}^2}{e^{-\lambda S} \lambda^2} & 0 \\ 0 & 0 & \frac{\lambda p_{NT} (1 - p_{NT})}{\lambda_{NT} (1 - e^{-\lambda S})} \end{pmatrix} \quad (8)$$

B.3.3 MEM2 Estimators

The equivalent equations for MEM2 (corresponding to the assumption that the probability of escape is the same for every species) with $p_T = p_{NT} = p$ can be obtained from the following LogLikelihood:

$$L(\lambda_T, \lambda_{NT}, p) = K - N_B S \lambda + (N - N_B) \log(1 - e^{-\lambda S}) + N_T \log(\lambda_T) \\ + N_{NT} \log(\lambda_{NT}) + (N_{NT} + N_T) \log(1 - p) - (N_T + N_{NT}) \log(\lambda) + N_E \log(p)$$

The first derivatives are then derived:

$$\frac{\partial L}{\partial \lambda_T} = -N_B S + \frac{(N - N_B) S e^{-\lambda S}}{1 - e^{-\lambda S}} + \frac{N_T}{\lambda_T} + \frac{N_T + N_{NT}}{\lambda}, \\ \frac{\partial L}{\partial \lambda_{NT}} = -N_B S + \frac{(N - N_B) S e^{-\lambda S}}{1 - e^{-\lambda S}} + \frac{N_T + N_{NT}}{\lambda_{NT}} + \frac{N - N_B}{\lambda}, \\ \frac{\partial L}{\partial p} = -\frac{N_T + N_{NT}}{1 - p} + \frac{N_E}{p}.$$

The maximum likelihood estimators are obtained by determining the roots of this set of equations.

The asymptotic covariance matrix is then given by:

$$COV_{MEM_2} = \frac{1}{N} \begin{pmatrix} \frac{\lambda_T \lambda_{NT}}{(1-p)(1-e^{-\lambda S})} + \frac{1-e^{-\lambda S}}{S^2 e^{-\lambda S}} \frac{\lambda_T^2}{\lambda^2} & -\frac{\lambda_T \lambda_{NT}}{(1-p)(1-e^{-\lambda S})} + \frac{1-e^{-\lambda S}}{S^2 e^{-\lambda S}} \frac{\lambda_T \lambda_{NT}}{\lambda^2} & 0 \\ -\frac{\lambda_T \lambda_{NT}}{(1-p)(1-e^{-\lambda S})} + \frac{1-e^{-\lambda S}}{S^2 e^{-\lambda S}} \frac{\lambda_T \lambda_{NT}}{\lambda^2} & \frac{\lambda_T \lambda_{NT}}{(1-p)(1-e^{-\lambda S})} + \frac{1-e^{-\lambda S}}{S^2 e^{-\lambda S}} \frac{\lambda_{NT}^2}{\lambda^2} & 0 \\ 0 & 0 & \frac{p(1-p)}{1-e^{-\lambda S}} \end{pmatrix} \quad (9)$$

C Some useful codes

C.1 JAGS code for MEM1

```
var
# data
# Data is a matrix with 4 columns and NData lines
# Col 1 corresponds to Nb, Col2 to N1, Col3 to N2, col4 to Ne
NData, P[NData], Data[NData,4], N[NData],
#variable requiring initialisation
lambda1, lambda2, p,
#variable without initialisation, deduced from the code
alpha[NData, 4];
```

```
model {
/* prior density */
p ~ dbeta(0.1,0.1) ;
```

```

lambda1 ~ dbeta(0.1,0.1);
lambda2 ~ dbeta(0.1,0.1);
lambda <- lambda1 + lambda2;

/* Model part */
for(j in 1:NData)
{
alpha[j,1] <- exp(-lambda * P[j]);
alpha[j,2] <- (1- exp(-lambda * P[j])) * lambda1 / lambda;
alpha[j,3] <- (1- exp(-lambda * P[j])) *lambda2/lambda *(1-p);
alpha[j,4] <- (1- exp(-lambda * P[j])) * lambda2 * p /lambda ;
Data[j,] ~ dmulti(alpha[j,], N[j]);
}
}

```

C.2 JAGS code for MEM2

```

var
# data
# Data is a matrix with 4 columns and NData lines
# Col 1 corresponds to Nb, Col2 to N1, Col3 to N2, col4 to Ne
NData, P[NData], Data[NData,4], N[NData],
# variable requiring initialisation
lambda1, lambda2, p,
# variable without initialisation, deduced from the code
alpha[NData, 4];

model {
/* prior density */
p ~ dbeta(0.1,0.1) ;
lambda1 ~ dbeta(0.1,0.1);
lambda2 ~ dbeta(0.1,0.1);
lambda <- lambda1 + lambda2;

/* Model part */
for(j in 1:NData)
{
alpha[j,1] <- exp(-lambda * P[j]);
alpha[j,2] <- (1- exp(-lambda * P[j])) * lambda1 * (1-p) / lambda;
alpha[j,3] <- (1- exp(-lambda * P[j])) *lambda2/lambda *(1-p);
alpha[j,4] <- (1- exp(-lambda * P[j])) * p ;
Data[j,] ~ dmulti(alpha[j,], N[j]);
}
}

```