**Expected Improvements in Precision when Integrating Opportunistic Close-Kin Mark-Recapture Data into Fisheries Stock Assessments**

Nicholas Fisch1

1Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada

Abstract

Close-Kin Mark-Recapture (CKMR) sampling, by providing information on abundance and survival rates (among other things), offers a promising new data source for fisheries stock assessments. To date and to my knowledge, most applications have focused on fitting CKMR models outside of stock assessment models. Fewer studies have directly integrated the CKMR data into the stock assessment model. Sample design in order to achieve a desired precision is somewhat straightforward in simple CKMR models however when integrated within a full stock assessment model with many other data sources, the value of the data (in terms of a reduction in uncertainty of model estimates) is less clear. Herein I demonstrate, using self-test simulations, the expected improvements in precision of derived quantities and estimated parameters within statistical catch-at-age models when opportunistic CKMR sampling is conducted and the data integrated within the assessment. By opportunistic CKMR sampling I mean to describe the genetic sampling of individuals that comprise the age composition data, such that increases in CKMR sampling would also increase the age composition samples (and vice versa). I examine the expected improvements (conditional on model being correct representation of the system) across three life history types (Cod-like, Flatfish-like, and Sardine-like) and different amounts of data available to the assessment, including the uncertainty and inclusion of an abundance index, and the sample size and time series length of CKMR and age composition samples.

abstract results

abstract discussion

Introduction

Close-Kin Mark-Recapture (CKMR, Bravington et al. 2016) offers a promising new source of information to fisheries stock assessments, which largely form the basis of commercial fisheries management throughout the world (Dichmont et al., 2016). Traditionally, stock assessments have been developed using routinely collected data from fisheries monitoring programs on total removals, age/length composition of the removals, and catch rates (Maunder and Punt, 2013). CKMR, given it allows for lethal sampling (e.g., sampling of the removals), may provide another data source to be routinely collected from commercial fisheries, contributing additional information on abundance and survival levels within fisheries stock assessments.

The theory of CKMR dates back to the early 2000s (Nielsen et al., 2001; Skaug, 2001) and since the publication of Bravington et al. (2016) has greatly increased in application within fisheries (see Casas and Saborido-Rey, (2023), for a suite of examples). The theory operates primarily by making inference on the number of parent-offspring and half-sibling pairs (POPs and HSPs) expected from a sample of a population. At its most basal level the theory posits that the expected number of POPs from a sample of a population is inversely related to the spawning abundance of that population. That is, the more POPs found in samples from a population, the smaller the population likely is, and vice versa. For HSPs, inference is based on an unobserved shared parent of the sibling pair, who must give birth to the older sibling and then survive to give birth to the younger sibling. Thus the expected number of HSPs in a sample is inversely related to both the adult mortality rates in the population and the spawning abundance in the year of each siblings birth. I avoid an in depth presentation of the theory as it is covered extensively in Bravington et al. (2016) and subsequent publications.

The lower limit of the precision of an abundance estimate from a CKMR model can be relatively straightforward to compute for the simplest case. For example, the lower limit of the coefficient of variation (CV) of an abundance estimate for a POP model with a single sampling event can be expressed as (Bravington and Grewe, 2007)

Where *m* denotes the total sample size of potential POPs examined (equally split among juveniles and adults), and the true abundance of adults in the population. Within an integrated fisheries assessment, the expected precision of estimates is less tractable and does not have a simple analytical solution, as it depends on a variety of different data sources and a complex, highly-parameterized non-linear model. However, the tool of self-test simulations can be employed to examine what might be expected in terms of precision of abundance estimates (and other desired quantities) when CKMR data are integrated into the assessment. According to CKMR theory, we expect that the information content of the CKMR data will depend on the size of the spawning abundance of the population and on the sample size collected. However, it is less clear how much specific additional information it will provide when combined with other data sources collected in fisheries assessment, such as age composition data (which, given that CKMR data are aged, will increase commensurate with increases in CKMR data). In addition, within a fisheries assessment, one might expect that even for cases where the population is very large and the CKMR sample size relatively small, zeroes in data (no matches) may help rule out the probability of a small population size.

In this study, I simulate a case where the age composition of the removals from populations are simultaneously observed for CKMR sampling. I refer to this opportunistic CKMR as no additional sampling effort is required other than genetic sampling of the age composition data that is already collected. This also has the effect of increasing the amount of composition data when the sample size of CKMR is increased. I therefore explore what one might expect in terms of improvements in model performance if CKMR data is collected and integrated within the stock assessment compared to identical assessments without CKMR data. I examine this comparison across three fish life history types (Cod-like, Flatfish-like, Sardine-like), in addition to across different amounts and quality of data such as the uncertainty in or the inclusion of an abundance index, and the sample size of CKMR and age-composition data.

My null-hypotheses are that…

Methods

The structure of the study was to have identical assessment models that were fit with and without CKMR data, so as to ascertain its value to the assessment. In the following I describe the population simulation operating model and the sampling model, including a separate component for the CKMR section.

Population simulation operating model

I chose to use similar population specifications to those used in Ono et al. (2015) and Hurtado-ferro et al. (2015) for simplicity and ease of comparison. The parameter values and equations governing the model are given in Tables 1 and 2, respectively. The population model runs for 100 years, with a fishing time series beginning in year 26 (years 26-100 experience fishing). The fishing time series was simulated as a two way trip with a linear increase from 0 in year 25 to a fully selected fishing mortality that was >*F*msy () and achieved 85% of maximum sustainable yield (MSY, at equilibrium) in year 85 of the time series, followed by a linear decrease to a fully selected fishing mortality that was <*F*msy () and also achieved 85% of MSY at equilibrium.

One hundred simulation iterations were run for each population model specification (Table 3), with stochasticity included using log-normal draws for recruitment deviations in each year. A sampling model then sampled from that population and fishery exploitation model to simulate the data collection process.

Sampling model

Annual landings were simulated using a log-normal distribution with a standard deviation (SD) of 0.05. A fishery index was simulated using a lognormal distribution where the SD was specified as a treatment within the simulation design, taking values of 0.25 or 0.5 (Table 3). The catch age-composition available for select years throughout the time series was simulated using a multinomial distribution. Age composition data was simulated from years 26, 36, 46, 51, 56, 61, 66, 71, and consistently from year 76 to 100. The number of samples prior to CKMR data collection began at *n*=30 in year 26 of the model (first year with fishing), increasing by 10 each decade until year 51, and then increasing by 10 every 5 years until year 76 where it was fixed at *n*=100 for the rest of the time series. This composition specification was similar to that used in Ono et al., (2015). Given CKMR data are aged samples, as CKMR data collection increased, as did the composition sample size, thus the composition sample size increased commensurate with CKMR data collection treatments (either 100, 1000, or 5000 sampled from the catch from the final 5, 10, or 20 years in the time series).

CKMR data

CKMR data were simulated from the terminal 5, 10, or 25 years of the time series (Table 3). Data were censused from the age composition for those years to make up the CKMR comparisons. That is, the age frequencies of the catch-at-age samples were those for which kinship comparisons were then made. Given the sampling year, the birth year of the older individual, and the birth year of the younger individual, the probability of being a POP or a HSP was calculated. The CKMR pair comparisons were aggregated to pairs with similar covariates (Table 4) and the number of POPs and HSPs for a given set of covariates was then simulated using a multinomial distribution. Note in the following that I do not distinguish “juveniles” and “adults” as the population is age structured with proportional maturity at age. Thus any individual whose age difference with another places them in a potentially mature age (at the year of birth of younger individual) has the potential to be a parent of the individual. I make comparisons between “younger” () and “older” () individuals in reference to both potential POPs and HSPs (rather than say, juveniles and adults). I limit comparisons to omit same cohort siblings (i.e., age difference must be greater than 0).

CKMR Simulator/Estimator

*Parent-offspring pairs*

Given the birth year of an older individual (), the birth year of a younger individual (), and the year the older individual was sampled (), the probability of the older individual being a parent-offspring pair with the younger individual is calculated as

which, describes the expected reproductive output of the potential parent in the year of younger individuals birth ( denotes age difference) divided by the total reproductive output in that year. A potential parent has to have been sampled after or on the year of younger individuals birth, because sampling is lethal in this case (as it is made up of the catch). Note that sex is not modeled and the total reproductive output is assumed to be made up of both males and females who contribute equally to spawning biomass. The 2 in the numerator reflects the fact that an individual has two parents. An alternative way of thinking about this probability, assuming a 50:50 sex ratio, is the probability that the older individual is female is 0.5 and then the probability that the female is a maternal-offspring pair (MOP) with the younger individual is Thus, adding the probability of a MOP () and a paternal-offspring pair gives

*Half-sibling pairs*

For HSPs, given the parent is not directly observed, a sum is taken over the probabilities of all the potential parents of the pair alive at the year of *i*’s birth (hence ). The first term within the summation describes the expected reproductive output of a potential parent (age ) over the total reproductive output in the year of the *i*’s birth. The addition of is to account for the abundance alive at *i*’s birth, and assuming a 50:50 sex ratio, 0.5 in the numerator and denominator cancel out. The second term is the probability that a potential parent survives from the time of *i*’s (older individual) birth to the time of *j*’s (younger individual) birth (i.e., through the year prior to *j*’s birth), assuming that spawning happens at the very start of the year. So a potential parent has to survive the year of *i*’s birth up to the age difference of the potential pair. Once the survival of the potential parent is accounted for, the third term in the equation is the probability of a parent-offspring pair between that potential parent and the younger individual (), where the 4 in the numerator in this case accounts for the fact that the parent is unobserved and could be a mother or a father (, Hillary et al. 2018).

Treatment Overview

Table 3 depicts the levels of each treatment type, which were implemented in a full factorial design to examine all combinations of treatments. Treatments included the species life history type, the sample size of CKMR and composition data, the length of time CKMR or increased composition data was collected, and the inclusion or sampling variance of the abundance index.

Estimation Model

The estimation/assessment models were identical in structure to the population operating model. They began in year 26 of the time series (first year with data). The fixed-effect parameters estimated within the model included the unfished recruitment, the recruitment SD natural mortality, two logistic selectivity parameters, catchability of the fishery index (if the index is included in assessment), and annual fully selected fishing mortality levels in each year (Table 1). Recruitment deviations in each year and those that make up the initial unfished abundance were treated as random effects. All other parameters were fixed at their true values from the operating model. The assessment models also fit the data using the same likelihoods used to generate data in the sampling model. In real-world situations, grandparent-grandchild pairs are difficult to distinguish genetically from HSPs (citation). In order to simulate a case where the chance of analyzing grandparent-grandchild pairs is limited, HSP comparisons within the multinomial likelihood were restricted to comparisons whose age difference was ≤2 times the age at 50% maturity for each life history (3, 4, and 3 for Cod-like, Flatfish-like, and Sardine-like, respectively).

Model Fitting

Models were fit in a mixed-effects context in Template Model Builder (TMB, Kristensen et al., 2016) with recruitment deviations specified as random effects and the remainder of the fixed effect parameters identified in Table 1. TMB calculates the marginal negative log-likelihood given the fixed effect parameters using the Laplace approximation to integrate over random effects. Fixed effect parameters are estimated via minimizing the marginal negative log-likelihood within the program R using the nlminb function. Random effects and derived quantities are then predicted using empirical Bayes (Kristensen et al., [2016](javascript:;)). Starting parameter values for each simulation iteration were chosen by sampling from a uniform distribution with the bounds specified as 50% below and 50% above the true parameter value. Standard errors of parameter estimates were extracted in addition to those of derived quantities using the generalized delta method built into TMB.

Comparison

Given the assessment models were correctly specified, it was expected that point estimates from assessments would be unbiased. I focused the comparison of the assessment models with and without CKMR data on the precision achieved between and within simulations. The precision between simulations was assessed by examining the standard deviation of the distribution of relative error over simulations, with relative error calculated as

Where denotes a point estimate from the assessment and the true value for that quantity in simulation iteration *s*. The precision within simulations was assessed by examining the estimated standard errors for quantities within an individual simulation iteration, and looking at the distribution of SEs over simulations. I chose to focus the comparison on spawning stock biomass over the time series, unfished recruitment, and natural mortality as they are of general interest in fisheries assessment and the CKMR data is thought to mainly inform abundance and survival (citation). For spawning stock biomass, precision within simulations was examined using coefficients of variation (CV) calculated as the standard error divided by the point estimate.

Results

Here describe that adding more data asymptotically led to unbiased estimates. And if there was bias, describe it.

Here described what happened in simulator regarding pairs.

All models were generally unbiased over simulations (Figure X).

Here describe the comparisons

Discussion

What happened with pairs. Bring in selectivity.

Supplemental file for exploitable abundance and selectivity.

Data accessibility

All code associated with analysis is available on *https://github.com/nickcfisch/CKMR\_Project*.

Acknowledgements

I would like to acknowledge and give sincere thanks to Eric Anderson and Paul Conn, whose workshop on close-kin mark-recapture at the 2022 annual meeting of the Wildlife Society introduced me to modeling associated with CKMR theory.

References